

# Cell

Volume 160  
Number 6

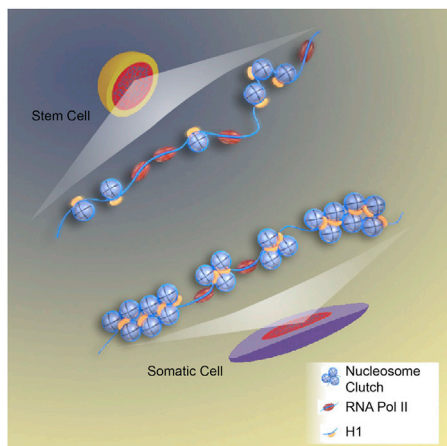
March 12, 2015

[www.cell.com](http://www.cell.com)



**Cold Tolerance in Rice**

**Dynamic and Organized Genome**



## Chromatin Fiber Irregulars

PAGE 1145

Using a quantitative super-resolution nanoscopy approach, Ricci et al. find that nucleosomes aggregate into clutches or groups of nucleosomes of differing size and density along chromatin fibers in vivo. These clutches represent a new level of chromatin architecture and speak to the functional state of individual cells as the clutch distribution reflects pluripotency potential for mouse ESCs.

## Lipid Sensing a TREMendous Loss in Alzheimer's

PAGE 1061

Recent genome-wide association studies have shown that a rare mutation of TREM2 correlates with a high risk of developing Alzheimer's disease. Wang et al. now find that TREM2 acts in microglia as a sensor for a wide array of lipids that are associated with  $\beta$ -amyloid accumulation and neuronal loss.

The mutant TREM2 attenuates microglial detection of damage-associated lipids, providing a mechanistic basis for the genetic association.

## Why Half Full Is Not Enough

PAGE 1072

Haploinsufficiency of genes encoding transcription factors can cause disease, but how? Theodoris et al. utilize endothelial cells derived from human iPSCs to show that heterozygous mutations in *NOTCH1* that cause cardiac defects disrupt the transcriptional and epigenetic response to shear stress, resulting in derepression of latent inflammatory gene networks. Computational predictions of the disrupted NOTCH1-dependent network reveal regulatory nodes for potential therapeutic intervention.

## A Fine RNA Balance for Neuronal Health

PAGE 1087

Pumilio1 is an RNA-binding protein that binds Ataxin1 mRNA and regulates its stability. Gennarino et al. report how haploinsufficiency of Pumilio1 results in an increase in endogenous Ataxin1 levels, leading to progressive motor dysfunction and degeneration of Purkinje cells, features typical of spinocerebellar ataxia type 1. These findings suggest that either haploinsufficiency of PUMILIO1 or duplication of ATAXIN1 could contribute to neurodegeneration in humans.

## Virus Kills Two Birds with One miRNA

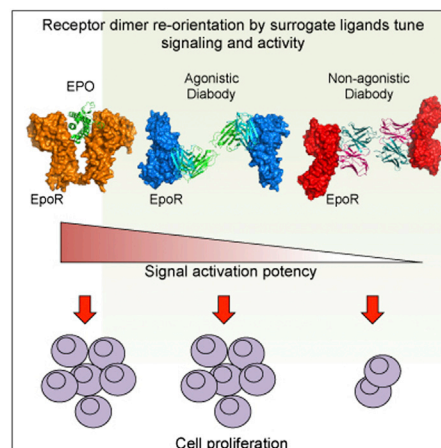
PAGE 1099

To replicate, hepatitis C virus (HCV) must bind the liver-specific tumor suppressor miRNA, miR-122. Luna et al. use sequencing and mathematical approaches to demonstrate that this interaction has a sponging effect that de-represses the liver mRNA targets of miR-122. Thus, in leveraging this miRNA as a replication factor, HCV simultaneously not only promotes its own propagation but also generates the necessary landscape to drive liver cancer.

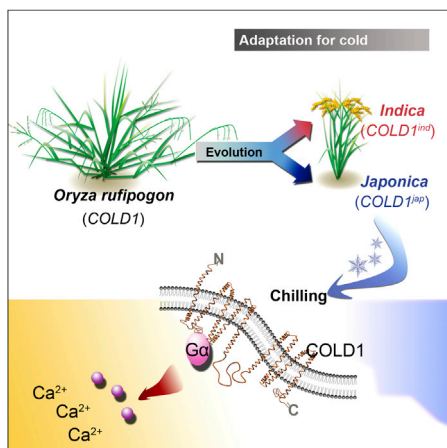
## Angling for the Right Signal

PAGE 1196

Most cell-surface receptors for cytokines and growth factors signal as dimers. Moraga et al. find that synthetic ligands called diabodies can be used to re-orient the geometry of receptor dimerization, thereby modulating the amplitude and nature of signal activation. Tuning receptor topology allows production of specific signal outputs, including correction of pathologic signals caused by oncogenic mutation of surface receptors.







## How Rice Moved North

PAGE 1209

Human selection has expanded the rice growth zone to regions with lower average temperature. Ma et al. identify that a single nucleotide polymorphism in the quantitative trait locus COLD1 underlies the adaptation to cold environment in japonica rice.

## This Message Will Self-Destruct

PAGE 1111

Codons within an mRNA not only dictate the sequence of the encoded protein but also convey information linked to the protein's expression level. Presnyak et al. now find that codon usage determines mRNA stability and impacts ribosome elongation rate. Proteins with related functions show similar patterns in codon content, demonstrating a new mechanism for coordinating functional gene expression through mRNA stability.

## circRNAs in EMT

PAGE 1125

Conn et al. show that the abundances of numerous circular RNAs (circRNA) are regulated during epithelial to mesenchymal transition, arguing for the functional involvement of circRNA in this process. A key regulatory protein for circRNA biogenesis in this context turns out to be the RNA binding protein Quaking that binds to intronic sites flanking circle-forming exons; indeed, insertion of Quaking binding sites is sufficient for circRNA generation.

## How Nucleosomes Unwind

PAGE 1135

Dynamic exposure of nucleosomal DNA plays key roles in many nuclear processes. Using single-molecule fluorescence-force spectroscopy, Ngo et al. address the relationship between DNA sequence and local nucleosome dynamics, showing that a nucleosome unravels asymmetrically under tension and that the direction of unwrapping is controlled by DNA flexibility.

## Movement without Motors

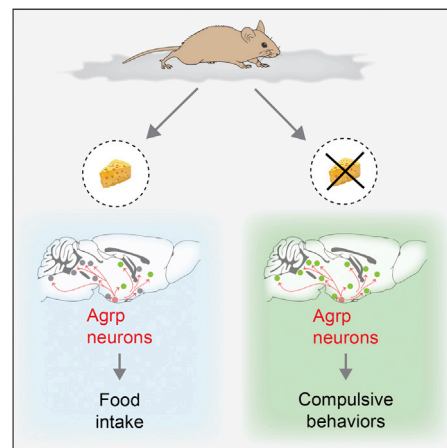
PAGE 1159

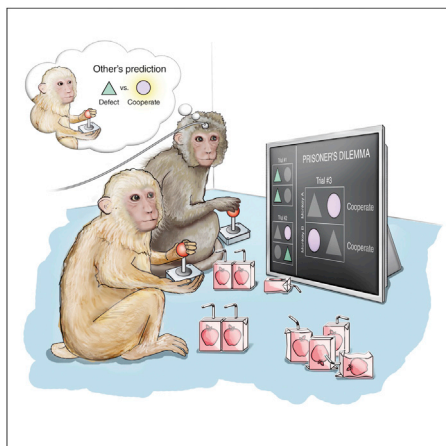
Mechanical forces driving cytoskeleton restructuring are attributed to the actions of molecular motors and the dynamics of cytoskeletal filaments, which consume chemical energy. Lansky et al. discover that mechanical force, comparable in magnitude to the force induced by microtubule motors, can be generated by passive diffusion of microtubule crosslinking proteins confined spaces. As confinement of diffusible crosslinkers is ubiquitous in cells, this mechanism is likely to be involved in many cellular processes.

## Feeding Circuit for More Than Food?

PAGE 1222

Activation of AgRP neurons in the hypothalamus increases appetite and feeding behavior if food is available. Dietrich et al. now find that, when food is not available, the activation of these neurons initiates compulsive behaviors such as foraging and marble burying. These observations unmask the relevance of primitive brain regions previously associated with energy homeostasis for complex behaviors beyond eating.





## Are You Thinking What I Think You're Thinking?

PAGE 1233

Successful social interchange relies on the ability to anticipate each other's intentions or actions. Williams and Haroush report the existence of cells in the cingulate cortex of primates that are able to anticipate the unknown intentions or state of mind of other individuals. These cells are critical for enacting cooperative social behavior. This framework might be relevant for understanding interpersonal, economic, and political decision-making process in humans.

## Transcriptional Debut Shapes Cell Cycle

PAGE 1169

The maternal-to-zygotic transition in early development is marked by slower cell-cycle progression and onset of de novo transcription; these have been thought to occur sequentially following dilution of a hypothetical limiting factor. Now, Blythe and Wieschaus report that, in fact, recruitment of RNA polymerase for zygotic transcription is what triggers the checkpoint for the cell-cycle delay

by interfering with DNA replication. Thus, an increasing load of zygotic transcription clashing with previously unimpeded S phases explains the cell-cycle remodeling.

## How Space Compresses Time

PAGE 1182

The point at which yeast re-enter the cell cycle is based on the history of pheromone exposure that had caused cell-cycle arrest. How can this memory be retained across a generation? Doncic et al. show that specific spatial organization of the G1/S switch components holds the key to past information.

## Screening Cancer Genes with CRISPR

PAGE 1246

Chen et al. use CRISPR/Cas9 to screen in vivo for mutations that drive tumor growth and metastasis. The hits include established tumor suppressors, as well as novel genes and microRNAs that are further validated. Cas9-based screening thus appears to be a robust method for systematically assaying mutant phenotypes in vivo.

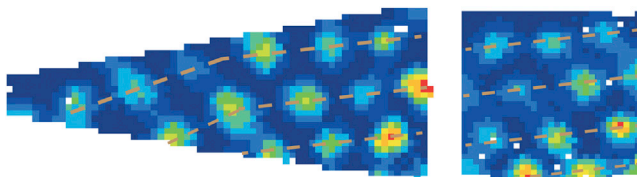


# Recalculating...

Knowing where we are and how to navigate through space is a skill we use on a daily basis, despite our increasing reliance on electronic GPS systems. Last year's Nobel Prize in Physiology or Medicine was jointly awarded to John O'Keefe and Edvard and May-Britt Moser for the remarkable discovery of cells in the brain responsible for spatial mapping (Kandel, 2014). A number of recent studies now provide even further insight into how our brain's internal GPS may work.

Grid cells are neurons in the cortex with unusual properties. When an animal is exploring a space, each individual grid cell fires when the animal is in multiple locations within that space, and these hot spots are spaced out in a hexagonal grid pattern, hence the name. Some models have suggested that grid cells provide an unchanging, universal matrix for measuring distances in space. However, two recent papers challenge this notion, demonstrating a clear influence of environmental geometry on grid pattern.

The orientation of grid patterns is thought to be anchored to reference frames in the environment that can be provided by distant cues. For example, when animals are in a symmetrical, circular space, grid cells use landmarks in the distance as an anchor. However, O'Keefe and colleagues found that when rats were placed in spaces with geometric features to which the animal might orient itself, such as a square, rotation of the environment resulted in rotation of grid axes (Krupic et al., 2015). This occurred even when landmarks in the distance stayed put, suggesting that when the local environment provides geometrical cues, the grid cells make use of them and reorient their firing patterns accordingly. But perhaps grid cells are still invariant in terms of their grid spacing? In fact, when rats were placed in a highly asymmetric environment, such as a trapezoid, their grid patterns bent and stretched to adapt to the geometry of the new environment and remained stably distorted. This indicates that grid cells deform their spatial activity patterns in a lasting way so that they fit the geometric features of the animal's local surroundings.

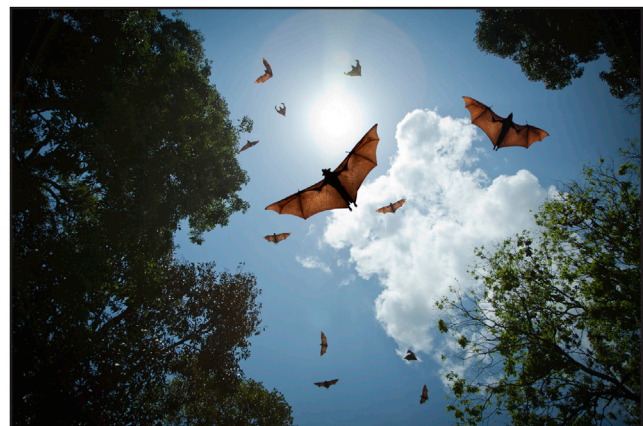


Grid patterns are symmetrical when rats are in a square environment (right) but are distorted in a trapezoidal environment (left). Image courtesy of J. Krupic and J. O'Keefe.

In another study, the Mosers and colleagues recorded from grid cells in rats exploring a square environment (Stensola et al., 2015). They found that the orientations of individual grid cell firing patterns were almost aligned with one of the walls of the square, but not quite—they were offset by a small degree, a finding also observed by O'Keefe and colleagues. Why might this be? It turns out that slight rotation minimizes

the symmetry between the grid pattern and the environmental geometry. If the grid patterns were perfectly aligned with the arena walls or offset by a wider angle, animals might get confused between different locations with the same geometric features, perhaps with one corner looking much like another. In addition, when they put rats into a square environment for the very first time, the grid patterns were not offset; instead, they were very closely aligned with one of the walls. This suggests that grid cells initially match up with the geometry of their new environment, but over time, the axes of grid patterns are rotated to optimize the animal's ability to determine its precise location.

Together, these two papers demonstrate that grid cells are more malleable than previously thought and can adapt their spatial firing patterns to the local environment the animal finds itself in. But what exactly is the role of these cells in spatial navigation? It has been postulated that grid cells use information about how far and in what direction an animal has moved in order to constantly update location. So it would seem reasonable to suppose that they receive input from neurons that can detect directionality. A study from Jeffrey Taube and colleagues now provides experimental evidence that grid cell function in rats requires input from head direction (HD) cells, which are found in part of the thalamus and represent the direction the animal is facing (Winter et al., 2015). When the authors inactivated these cells by local injection of lidocaine, which blocks sodium channel function, or by severely lesioning the brain region containing these cells, grid-like firing in the cortex disappeared.



Bats navigating through three-dimensional space. Image from iStock.com/peters99

The above studies were carried out in rats exploring a two dimensional environment, but what about animals that can fly? How do they constantly evaluate their three-dimensional position and heading as they navigate through space? A recent paper from Nachum Ulanovsky and colleagues shows that bats not only have HD cells, but several different flavors of them (Finkelstein et al., 2015). While some are direction

selective in the horizontal plane, others are tuned to pitch or roll, and still others respond two or three dimensionally. Remarkably, when bats flip to an inverted hanging position, the direction selectivity of horizontally sensitive neurons is shifted by  $180^\circ$  by the time they're upside-down. This means that instead of having to activate a new population of neurons during such acrobatic moves, bats use the same sets of horizontally selective neurons to guide them stably through the aerial maneuver. But if this is the case, then how do bats know if they're upside down or upright? It turns out that the population of cells tuned to the vertical orientation cover the whole  $360^\circ$  range for pitch, so together with the horizontally coded cells, these allow the bats to tell their precise three-dimensional orientation. From this data, the authors propose that bats use a donut-shaped coordinate system to represent head direction. An advantage of using toroid rather than spherical coordinates is that abrupt discontinuities are avoided when pitch changes dramatically, allowing smooth representation of three-dimensional position.

Since the study from Taube and colleagues indicates that HD cells provide input to grid cells, this opens up the possibility that grid cells also code in three-dimensional space. Another intriguing question is whether non-flying mammals that navigate in complex three-dimensional environments use a similar system. For example, when your cat is engaging in its nightly acrobatic antics, is it making use of a complex, three-dimensional neural compass? What about humans, especially gymnasts and ski jumpers who are able to perform remarkable, gravity-defying feats? There's evidence that even in rats, which we normally think of as rather earth-bound creatures, navigational neurons are able to code for three-dimensional space (Hayman et al., 2011), suggesting that at least some version of a three-dimensional compass may be widespread among mammals. It will be interesting to learn whether experience and training are able to shape the function of such a compass.

## REFERENCES

- Finkelstein, A., Derdikman, D., Rubin, A., Foerster, J.N., Las, L., and Ulanovsky, N. (2015). *Nature* 517, 159–164.
- Hayman, R., Verriotes, M.A., Jovalekic, A., Fenton, A.A., and Jeffery, K.J. (2011). *Nat. Neurosci.* 14, 1182–1188.
- Kandel, E. (2014). *Cell* 159, 1239–1242.
- Krupic, J., Bauza, M., Burton, S., Barry, C., and O'Keefe, J. (2015). *Nature* 518, 232–235.
- Stensola, T., Stensola, H., Moser, M.B., and Moser, E.I. (2015). *Nature* 518, 207–212.
- Winter, S.S., Clark, B.J., and Taube, J.S. (2015). *Science* 347, 870–874.

**Cindy Lu**



# A Molecular Take on Aesop's *The Oak and the Reeds*

Corentin J. Moevus<sup>1</sup> and Eric C. Greene<sup>2,3,\*</sup>

<sup>1</sup>Department of Pathology and Cell Biology

<sup>2</sup>Department of Biochemistry and Molecular Biophysics

<sup>3</sup>Howard Hughes Medical Institute

Columbia University, 650 West 168th Street, New York, NY 10032, USA

\*Correspondence: [ecg2108@columbia.edu](mailto:ecg2108@columbia.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.036>

**Ngo et al. use single-molecule methods to show that DNA can be more readily displaced from one side of a nucleosome relative to the other side. This unexpected mechanical asymmetry may offer a path of least resistance, allowing RNA polymerases to traverse nucleosomes if they approach from the correct direction.**

A very stiff nucleosome was uprooted by the RNA polymerase and the freed histones diffused around the genome. They wandered around some flexible nucleosomes, which they thus addressed: “I wonder how you, who established on so bendable and weak DNA, are not entirely evicted by the polymerases.” They replied, “You fight and contend with your DNA, and consequently you are destroyed; while our sequences on the contrary bend around every contact site, and therefore we remain unbroken.”

—Molecular Aesop

In Aesop's famous fable *The Oak and the Reeds*, a proud and stiff oak is uprooted by strong winds, whereas the humble and flexible reeds bend and survive the storm. Little did Aesop know that his wisdom would hold at the molecular level. In this issue of *Cell*, Ngo and colleagues (Ngo et al., 2015) demonstrate that differences in DNA flexibility between the two halves of a nucleosome can lead to a strong asymmetric behavior, the more flexible half being more stable (Figure 1).

Eukaryotic genomes are organized into chromatin, the smallest repeating unit of which is the nucleosome: a symmetric structure composed of ~147 base pairs (bp) of DNA wrapped 1.75 times around an octamer of histone proteins. Nucleosomes cover up to 90% of the genome and inhibit access to the underlying DNA

by steric hindrance. How then can processes such as DNA repair, replication, or transcription happen in the context of chromatin?

Proteins can access nucleosomal DNA either passively, via spontaneous site exposure, or actively, via chromatin remodeling. Jonathan Widom's lab pioneered the work on spontaneous site exposure. Each of the 14 contact sites between the histone octamer and DNA can detach transiently, freeing around 10 bp of DNA per site for other proteins to capture. The microscopic dissociation constants at specific locations of the nucleosome were obtained by FRET (fluorescence resonance energy transfer) measurements (Li et al., 2005; Tomschik et al., 2005), demonstrating the model. This line of work led to the idea that proteins such as transcription factors can first bind to nucleosomal DNA passively to subsequently be used as platforms for the recruitment of chromatin remodelers that enhance access to the obstructed genetic information.

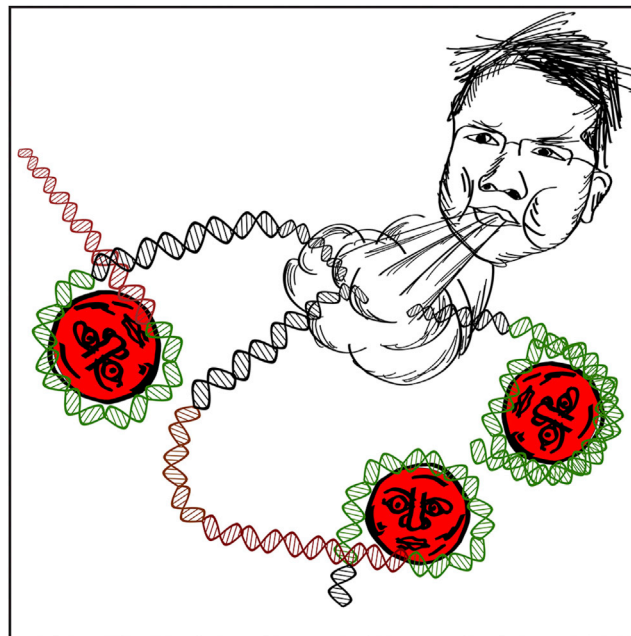
As the study of nucleosome dynamics switched toward single-molecule methods (Killian et al., 2012), force measurement techniques such as optical tweezers were used to mimic (Hall et al., 2009) or measure (Hodges et al., 2009) a polymerase accessing nucleosomal DNA. These studies permitted experimental access to both the nucleosome and the polymerase during an encounter. The former presents two main barriers against force before being evicted, whereas the latter stumbles and backtracks on nucleosomal

DNA. However, force measurements usually lack the three-dimensional spatial information that can be provided by FRET: it is difficult to link the observed behavior with a specific structure of the nucleosome. Which specific regions of the nucleosome give rise to these two main barriers? What specifically happens to the sub-structures of the nucleosome? FRET measurements have the potential to provide this information, but a standard FRET experiment does not allow concurrent measures of mechanical stability. To solve this problem, Ngo and colleagues merge the two techniques, single-molecule FRET and optical tweezers, to observe precisely which DNA-histone interactions get disrupted when tension is applied to the nucleosome.

Using this powerful approach, they confirm the long acknowledged link between nucleosome stability and DNA flexibility (Cloutier and Widom, 2005): the stiffer the DNA sequence, the less stable the nucleosome. They furthered this observation at the sub-nucleosome level. The two symmetric halves of a nucleosome, spanning around the dyad, comprise different DNA sequences, unless the DNA is palindromic. Consequently, one half of the nucleosome can be more flexible and therefore more stable than the other half. This is indeed what they observe, and it explains their key finding: unwrapping of a nucleosome under force can be asymmetric. The stiffer side will unwrap first, and as it happens, the more flexible side will be stabilized. This finding is of particular interest

because it implies that a nucleosome can be more easily bypassed from one side than the other. Interestingly, this has been observed in vitro during transcription of nucleosomal DNA (Bondarenko et al., 2006). This is another piece of molecular wisdom: when facing an obstacle, be sure to approach it from the correct direction!

As the authors point out, the asymmetry in the mechanical stability of nucleosomes can bring about an unanticipated level of gene regulation. By being a stronger barrier in one direction compared to the other, a nucleosome can permit transcription to happen only in a given direction. This would be an effective way to prevent antisense transcription, and one could readily envision how the directional stability of an entire array of nucleosomes spanning a gene could amplify this effect. However, the Widom 601 sequence used by Ngo et al. emerges from in vitro selection for strong nucleosome binding sequences (Lowary and Widom, 1998) and does not normally exist in living organisms. Even though the Widom 601 sequence is by far the strongest known nucleosome positioning sequence, it still cannot prevent transcription in either orientation in vivo (Perales



**Figure 1. Asymmetric Response of Nucleosomes under Tension**

Under force (here, depicted as wind), nucleosomes display asymmetric dynamics. The stiffer half of the nucleosomal DNA (brown) will be disrupted first, whereas the more flexible one (green) will remain bound. For a DNA sequence of high flexibility, the nucleosome will remain bound at higher forces that would otherwise evict nucleosomes from stiffer DNA sequences.

et al., 2011). This indicates that polymerases can, perhaps by recruiting a chromatin remodeler, bypass even the strongest known nucleosome barrier.

In summary, Ngo and colleagues manage to reveal the unexpected asymmetry of nucleosome stability by merging two widely successful techniques in this field: single-molecule FRET and optical tweezers. Their work nicely fits with

existing literature and should prompt interest in future studies to help determine whether nature has co-opted this remarkable structural and mechanical asymmetry as a potential means for regulating gene expression.

## REFERENCES

- Bondarenko, V.A., Steele, L.M., Ujvári, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). *Mol. Cell* 24, 469–479.
- Cloutier, T.E., and Widom, J. (2005). *Proc. Natl. Acad. Sci. USA* 102, 3645–3650.
- Hall, M.A., Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T., and Wang, M.D. (2009). *Nat. Struct. Mol. Biol.* 16, 124–129.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). *Science* 325, 626–628.
- Killian, J.L., Li, M., Sheinin, M.Y., and Wang, M.D. (2012). *Curr. Opin. Struct. Biol.* 22, 80–87.
- Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). *Nat. Struct. Mol. Biol.* 12, 46–53.
- Lowary, P.T., and Widom, J. (1998). *J. Mol. Biol.* 276, 19–42.
- Ngo, T.T.M., Zhang, Q., Zhou, R., Yodh, J.G., and Ha, T. (2015). *Cell* 160, this issue, 1135–1144.
- Perales, R., Zhang, L., and Bentley, D. (2011). *Mol. Cell. Biol.* 31, 3485–3496.
- Tomschik, M., Zheng, H., van Holde, K., Zlatanova, J., and Leuba, S.H. (2005). *Proc. Natl. Acad. Sci. USA* 102, 3278–3283.



# Mitosis, Diffusible Crosslinkers, and the Ideal Gas Law

David J. Odde<sup>1,\*</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Minnesota, Minneapolis, MN 55455, USA

\*Correspondence: [odde002@umn.edu](mailto:odde002@umn.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.048>

**During mitosis, molecular motors hydrolyze ATP to generate sliding forces between adjacent microtubules and form the bipolar mitotic spindle. Lansky et al. now show that the diffusible microtubule crosslinker Ase1p can generate sliding forces between adjacent microtubules, and it does so without ATP hydrolysis.**

The mitotic spindle is organized by an ensemble of molecular motors that hydrolyze ATP to actively transport microtubules. For example, the kinesin-5 family molecular motors (Cin8/Eg5/Kif11) generate sliding forces between anti-parallel microtubules to push spindle poles apart, establish the metaphase bipolar spindle, and ultimately physically separate replicated genomes (Subramanian and Kapoor, 2012). These motors are resisted by passive diffusible crosslinkers, such as Ase1/PRC1/Map65, that have previously been viewed as mere frictional elements (Braun et al., 2011; Pringle et al., 2013). Since friction always acts against the direction of relative movement, the Ase1p-mediated frictional force in this overdamped system would then be predicted to drop to zero once an applied force was removed. In this issue, Lansky et al. show that this prediction is not observed, but rather that Ase1p drives microtubule sliding to maximize overlap in the absence of any applied force or ATP (Lansky et al., 2015).

To investigate force generation mediated by Ase1p crosslinkers, Lansky et al. used an in vitro experiment with purified Ase1p-GFP and red fluorescent microtubules. One “template” microtubule was firmly attached to a coverslip, and then a second microtubule was crosslinked to the template via Ase1p and the ensemble imaged via total internal reflection fluorescence microscopy. The ensemble was then subjected to a variety of forces, including hydrodynamic flow, optical tweezers, and molecular motors, that displaced the microtubules relative to

each other, thus reducing the overlap region, as depicted in Figure 1A. As shown previously, continued force application will eventually slide the two apart completely (Braun et al., 2011). However, when the applied force was suddenly removed before all overlap was lost, a strange thing occurred: the second microtubule slowly slid backward to regain the lost overlap between the two microtubules. On the nanometer scale of the molecules, the observed displacements were large covering micrometers. The equivalent macroscopic experiment might be dragging a pencil across a table until it hangs over the edge of the desk, then letting go and seeing the pencil creep back onto the desk. Where does the force come from when there is no ATP or micrometers-long spring to drive the recovery of the overlap? Surprisingly, the familiar ideal gas law,  $PV = nRT$ , governs the system.

Unlike the pencil experiment, the microtubule experiment is strongly influenced by thermal forces. As a result, Ase1p can explore a variety of positions within the overlap. As the overlap increases, more positions become available to the Ase1p, as shown in Figure 1B. Thus, the greatest number of positions is accessed when overlap is maximal. Since these positions are energetically equivalent, the most probable state of the system is maximal overlap. If one were to apply a force, this would limit the number of accessible states and compress Ase1p into a smaller overlap region. This is the same physics of an ideal gas, as expressed in the ideal gas law. In this linear system, the ideal gas law can be

written  $FL = nk_B T$ , where  $F$  is the force,  $L$  is the overlap length,  $n$  is the number of crosslinkers,  $k_B$  is Boltzmann’s constant, and  $T$  is the absolute temperature. As the overlap decreases, the force builds as  $F \sim 1/L$ , which is observed experimentally.

This is a beautiful experimental demonstration of entropy maximization at work. The entropy,  $S$ , for any state of the system is given by

$$S = k_B \ln W$$

where  $W$  is the multiplicity of the state given by

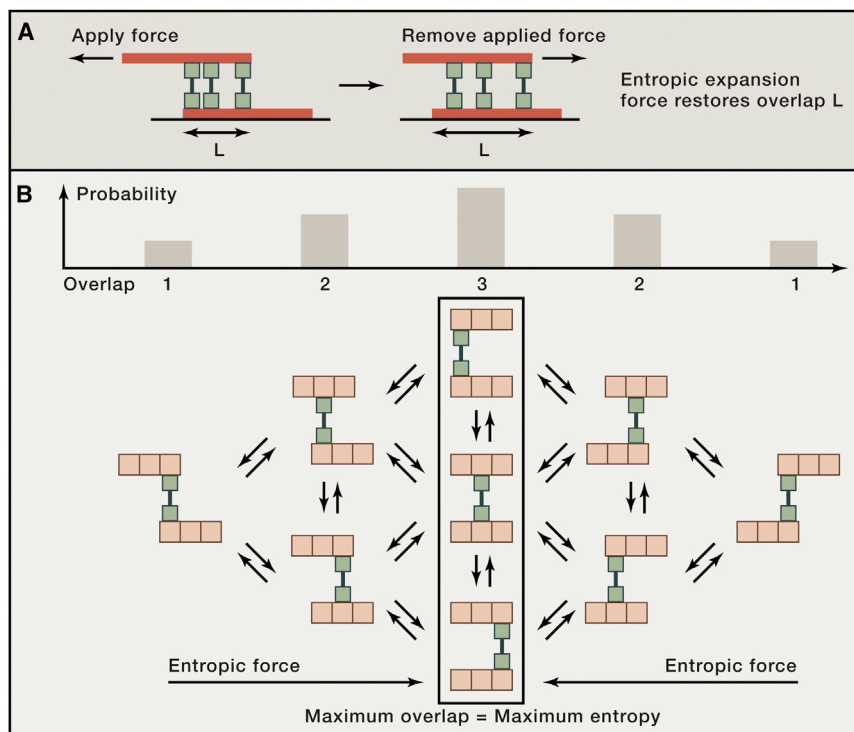
$$W = \frac{M!}{N!(M-N)!}$$

where  $M$  is the number of configurations and  $N$  is the number of molecules. When  $S$  is maximal, the Gibbs free energy,  $G$ , is minimal (assuming no net change in the number of crosslinking bonds). The more probable a state is, the greater the entropy of that state. In the case of microtubule sliding, the more overlap between the microtubules, the more possible configurations there are that achieve that state, as illustrated in Figure 1B. For a single diffusing molecule,  $N = 1$ , and

$$W = \frac{M!}{1!(M-1)!} = M$$

For example, for overlap = 1, there is only one possible configuration of the single crosslinker ( $W = 1$ ). Thus, for overlap = 1, the entropy is

$$S = k_B \ln(1) = 0$$



**Figure 1. Diffusible Crosslinkers Drive an Entropic Expansion Force to Maximize Overlap between Adjacent Microtubules**

(A) Microtubules (red) are crosslinked with Ase1p (green), which can diffuse along the microtubule surfaces. Ase1p exerts passive frictional resistance to applied forces that displace one microtubule relative to the other. Lansky et al. show that when the force is relieved, the microtubule slides back to re-establish maximal overlap,  $L$ , between the microtubules. Like a compressed ideal gas, the expansion of Ase1p along the lattice creates the restoring force.

(B) Origin of the entropic expansion force. In this example, two microtubules of length 3 are crosslinked by one Ase1p. Since there is only 1 way to achieve the left-most configuration, it is less probable than the overlap = 2 (2 possible configurations) and overlap = 3 (3 possible configurations) cases, and equally probable to the rightmost overlap = 1 case. Therefore, the most overlapped (overlap = 3) state is the most probable, and so the entropy is maximal. This creates a driving force toward maximal overlap, as observed by Lansky et al.

For overlap = 2, there are two possible configurations ( $W = 2$ ), and so

$$S = k_B \ln(2)$$

and for the most overlapped state (overlap = 3), there are 3 possible configurations ( $W = 3$ ), and so the entropy is

$$S = k_B \ln(3)$$

So we see that the entropy is maximal for the most overlapped state, and driving the system away from this state requires an applied force.

In terms of free energy,  $\Delta G = -T\Delta S$ , the biggest change in Figure 1B occurs when overlap increases from 1 to 2, which is  $\Delta G = -T\Delta S = -\ln(2)k_B T = -0.69k_B T$ . Since the force,  $F = -\Delta G/\delta$ , where  $\delta$  is

the distance over which the energy change occurs, we can then estimate the entropic expansion force. Assuming a step size of  $\delta = 4$  nm, which is the size of a tubulin monomer, and an energy unit conversion of  $1 k_B T = 4.28$  pN-nm, then the entropic force is  $F = (0.69 k_B T) (4.28 \text{ pN-nm}/k_B T) / (4 \text{ nm}) = 0.7$  pN, comparable to the force exerted by a molecular motor. Adding more crosslinkers would cause the force to increase proportionately, which Lansky et al. also demonstrate experimentally. Thus, the authors view the crosslinkers as exerting an “entropic expansion force” that acts to maximize the overlap between the two microtubules.

The entropic force is distinct from molecular motor forces in that it does

not require ATP hydrolysis. It is also distinct from the microtubule depolymerization force, which drives kinetochore poleward movements in mitosis, a.k.a. the Hill sleeve mechanism (Hill, 1985; Powers et al., 2009). More generally, the importance of entropic forces is already appreciated in determining disordered protein acid structure, and in the packaging of viral genomes (Bustamante et al., 1994). Lansky et al. now reveal another entropy-driven force generating mechanism based on diffusible crosslinkers driving increased overlap between two adjacent self-assembled linear polymers.

So what do these findings mean for cells? It seems strange that Ase1p has the ability *in vivo* to enhance pole separation (Syrovatkina et al., 2013), but this counterintuitive effect is perhaps explained by Ase1p’s bundling activity. This activity makes kinesin-5 more efficient as recently reported for the minus end-directed motor Kar3-Cik1 (Hepperla et al., 2014). What it does mean is that the pole-separating kinesin-5 motors may be working harder than we previously thought because they must overcome the extra entropic force that acts in the background to collapse the spindle. In this light, the entropic force may therefore help stabilize the spindle midzone in late mitosis. Beyond microtubules, Lansky et al. speculate that the same principles might drive sliding of actin filaments in cytokinesis due to diffusible crosslinking by myosin II, for example, rather than by its motor activity. At the cellular scale, it seems possible that diffusible crosslinkers that bridge between adjacent cells, such as cadherins, could also exert an entropic force that by itself would act to maximize contact area between adherent cells. In general, the ideal gas law is likely embedded in the background of a multitude of thermally driven cellular processes, exerting forces in the constant search for maximal entropy.

## REFERENCES

- Braun, M., Lansky, Z., Fink, G., Ruhnaw, F., Diez, S., and Janson, M.E. (2011). Nat. Cell Biol. 13, 1259–1264.
- Bustamante, C., Marko, J.F., Siggia, E.D., and Smith, S. (1994). Science 265, 1599–1600.



Hepperla, A.J., Willey, P.T., Coombes, C.E., Schuster, B.M., Gerami-Nejad, M., McClellan, M., Mukherjee, S., Fox, J., Winey, M., Odde, D.J., et al. (2014). *Dev. Cell* 31, 61–72.

Hill, T.L. (1985). *Proc. Natl. Acad. Sci. USA* 82, 4404–4408.

Lansky, Z., Braun, M., Ludecke, A., Schlierf, M., ten Wold, P.R., Janson, M.E., and Diez, S. (2015). *Cell* 160, this issue, 1159–1168.

Powers, A.F., Franck, A.D., Gestaut, D.R., Cooper, J., Graczyk, B., Wei, R.R., Wordeman, L., Davis, T.N., and Asbury, C.L. (2009). *Cell* 136, 865–875.

Pringle, J., Muthukumar, A., Tan, A., Crankshaw, L., Conway, L., and Ross, J.L. (2013). *J. Phys. Condens. Matter* 25, 374103.

Subramanian, R., and Kapoor, T.M. (2012). *Dev. Cell* 23, 874–885.

Syrovatkina, V., Fu, C., and Tran, P.T. (2013). *Curr. Biol.* 23, 2423–2429.

# Transcription Gets to the Checkpoint

John D. Laver<sup>1</sup> and Howard D. Lipshitz<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, ON M5S1A8, Canada

\*Correspondence: [howard.lipshitz@utoronto.ca](mailto:howard.lipshitz@utoronto.ca)  
<http://dx.doi.org/10.1016/j.cell.2015.02.051>

**The rapid cell proliferation characteristic of early animal embryos is accomplished with an abbreviated cell cycle and no DNA replication checkpoint. Blythe and Wieschaus provide evidence that nascent zygotic transcription precedes—and may trigger—this checkpoint at the midblastula transition.**

During the cell cycle, the DNA replication checkpoint pauses entry into M phase until replication is complete. Activation of this checkpoint is essential in early embryos of many animals. In *Drosophila*, for example, a deficient checkpoint results in severe mitotic defects and death (Sibon et al., 1997). Although the importance of the checkpoint is clear, how and why it is activated in early embryos is less so. In this issue of *Cell*, Blythe and Wieschaus (2015) present evidence that checkpoint activation in *Drosophila* is triggered by the onset of zygotic transcription (Figure 1).

The earliest phase of development in *Drosophila* consists of 13 rapid, synchronous nuclear cycles (NCs)—composed only of S and M phases—directed by maternally supplied mRNAs and proteins. As development proceeds, maternal products are degraded and the zygotic genome is activated, a process known as the maternal-to-zygotic transition (MZT). Concurrently, gradual lengthening of the NCs culminates in the introduction of gap phases and cellularization of the blastoderm during NC14, an event known as the midblastula transition (MBT). These processes depend on a functional replication checkpoint.

A long-standing model posits that, with increasing nucleocytoplasmic ratio,

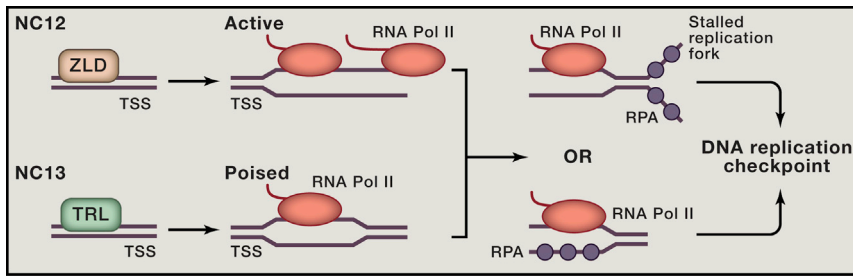
essential maternal replication factors are titrated, resulting in replication stress and checkpoint activation (Sibon et al., 1997). In a series of ingenious experiments, Blythe and Wieschaus (2015) use compound chromosomes to alter the total DNA content of the embryo or to modulate the amount of transcriptionally active DNA in embryos with the same total DNA content. By precisely measuring the length of NC13 as a proxy for the extent of checkpoint activation, they demonstrate that this activation correlates best not with total embryonic DNA content but with the amount of transcriptionally engaged DNA, leading to the hypothesis that checkpoint activation is a consequence of the onset of zygotic transcription.

To test this model, Blythe and Wieschaus (2015) perform RNA polymerase II (Pol II) chromatin immunoprecipitation sequencing (ChIP-seq) on carefully staged embryos to accurately define changes in transcriptional activity in NC12, NC13, and NC14. While hundreds of genes are already occupied and undergoing transcription at NC12, NC13 marks the large-scale recruitment of Pol II, largely in a “poised” state, to the transcriptional start sites of thousands of additional genes, which is consistent with the results of an earlier study (Chen et al., 2013). Importantly, these early

phases of global zygotic genome activation are largely unaffected in checkpoint mutants, implying that transcription precedes and occurs independently of checkpoint-mediated NC lengthening.

To monitor replication stress at the molecular level, Blythe and Wieschaus (2015) next use fluorescently labeled RPA70, which binds to sites of single-stranded DNA generated upon replication stalling, leading to checkpoint activation. They demonstrate a striking correlation between RPA70-bound and Pol II-occupied DNA, which is consistent with the hypothesis that sites of transcriptionally engaged DNA are sources of replication stress. This interpretation is complicated by the fact that, in budding yeast, RPA70 is generally associated with sites of active transcription independent of replication (Sikorski et al., 2011), so it remains possible that the correlation reflects not sites of replication stalling but a role for the RPA complex in transcription. Indeed, Blythe and Wieschaus (2015) speculate that RPA may directly link transcription to the checkpoint independent of replication stress. Assessing additional and highly specific markers of replication stress, such as phosphorylated RPA30, may be illuminating.

The most compelling evidence for a transcription-induced checkpoint model



**Figure 1. Blythe and Wieschaus Propose that the Onset of Zygotic Transcription Triggers the Activation of the DNA Replication Checkpoint at the Midblastula Transition in *Drosophila***

In this model, zygotic transcription is activated by transcription factors such as Zelda (ZLD), which binds upstream of the transcription start site (TSS) and promotes active transcription starting at nuclear cycle (NC) 12, and Trithorax-like (TRL), which recruits Pol II in a “poised” state at NC13. Actively transcribing and/or poised Pol II leads to recruitment of the RPA complex, either by causing replication stalling (top right) or through direct recruitment to sites of transcription (bottom right). RPA then activates the replication checkpoint and the associated cell-cycle remodeling characteristic of the MBT.

comes from their finding that decreasing the amount of Pol II-bound DNA suppresses the mitotic catastrophe caused by mutations in the *mei-41/ATR* checkpoint gene. First, suppression is achieved with mutations in the transcription factor Zelda (Vielfaltig), which is required for the early phase of zygotic transcription and which the authors confirm by Pol II ChIP-seq, mediates active transcription starting at NC12. This is consistent with an earlier observation that premature zygotic transcription leads to Zelda-dependent premature checkpoint activation (Sung et al., 2013). Second, mutations in the transcription factor Trithorax-like (GAGA Factor), which has been predicted to have a role in the establishment of poised Pol II at NC13 and 14 (Chen et al., 2013), also suppress *mei-41*. Thus, either a reduction in the amount of active transcription or a reduction in poised Pol II can partially mitigate the absence of a replication checkpoint.

In the future, it will be important to determine the relative contribution of active versus poised Pol II to checkpoint activation. In addition to their genetic suppression experiments, other data presented by Blythe and Wieschaus (2015) suggest a joint role. In the absence of Zelda, both Pol II and RPA70 are reduced at sites of active transcription rather than at poised sites, supporting a role for active transcription. However, the lengthening of NC13 in different compound-chromosome combinations correlates primarily with the recruitment of poised Pol II. Likewise, treatment with  $\alpha$ -amanitin does not suppress the *mei-41* mitotic

catastrophe, suggesting that the checkpoint trigger precedes active transcriptional elongation.

Whatever the relative roles of poising and active transcription, the transcription-induced checkpoint hypothesis provides an intriguing link between the hand-over of developmental control from the maternal to the zygotic genome and the concurrent changes in the cell cycle that occur during early development. If the onset of zygotic transcription triggers the replication checkpoint at the MBT, what triggers the onset of zygotic transcription? Zygotic transcripts fall into two classes; the minority depend on the nucleocytoplasmic ratio for transcription, and the majority are transcriptionally activated independent of this ratio (Lu et al., 2009). The former could be activated by checkpoint-independent increases in cell-cycle length mediated by titration of maternal factors such as Cyclin B (Edgar et al., 1994), whereas the latter likely depend on a maternal timer acting independently of cell-cycle changes. One candidate for such a timer is the RNA-binding protein Smaug, which directs degradation of maternal transcripts during the MZT (Tadros et al., 2007) and is required for high-level expression of the zygotic genome, as well as checkpoint activation (Benoit et al., 2009). Smaug levels gradually increase in early embryos, peaking at the MBT, and alterations in the amount of Smaug affect the timing of the MZT and MBT (Benoit et al., 2009). Smaug might promote the clearance of maternally supplied transcriptional repressors, thus permitting activation of the zygotic

genome. A second such timer might be Zelda itself, whose accumulation could time the activation of early zygotic transcription. Together, Smaug and Zelda would, respectively, be permissive and instructive timers of zygotic genome activation. Further studies will be required to tease out the roles of these and other repressors and activators of zygotic transcription and to clarify the role of the different classes of zygotic transcripts in activating the checkpoint.

Finally, it will be interesting to determine whether the onset of zygotic transcription has a role in triggering the replication checkpoint at the MBT in other species. In *Xenopus*, overexpression of a subset of replication factors extends the early, rapid embryonic cell divisions (Collart et al., 2013), supporting the model that maternally supplied replication factors are titrated, resulting in replication stress and checkpoint activation. Although Blythe and Wieschaus (2015) provide evidence that replication factors may not be limiting in *Drosophila*, the two models need not be mutually exclusive—multiple factors could help pull the MBT checkpoint trigger.

## REFERENCES

- Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). Development 136, 923–932.
- Blythe, S.A., and Wieschaus, E.F. (2015). Cell 160, this issue, 1169–1181.
- Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., and Zeitlinger, J. (2013). eLife 2, e00861.
- Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Science 341, 893–896.
- Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P., and O’Farrell, P.H. (1994). Genes Dev. 8, 440–452.
- Lu, X., Li, J.M., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2009). Development 136, 2101–2110.
- Sibon, O.C., Stevenson, V.A., and Theurkauf, W.E. (1997). Nature 388, 93–97.
- Sikorski, T.W., Ficarro, S.B., Holik, J., Kim, T., Rando, O.J., Marto, J.A., and Buratowski, S. (2011). Mol. Cell 44, 397–409.
- Sung, H.W., Spangenberg, S., Vogt, N., and Großhans, J. (2013). Curr. Biol. 23, 133–138.
- Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). Dev. Cell 12, 143–155.

# Cold Tolerance Encoded in One SNP

Prabha Manishankar<sup>1</sup> and Jörg Kudla<sup>1,\*</sup>

<sup>1</sup>Institut für Biologie und Biotechnologie der Pflanzen, Universität Münster, Schlossplatz 7, 48149 Münster, Germany

\*Correspondence: [jkudla@uni-muenster.de](mailto:jkudla@uni-muenster.de)  
<http://dx.doi.org/10.1016/j.cell.2015.02.037>

**Cold tolerance fundamentally affects world crop harvest. Ma et al. now identify a single-nucleotide polymorphism in a gene called *COLD1* that confers cold tolerance in *japonica* rice. This study reveals important insights into agronomical traits that are essential for human nutrition.**

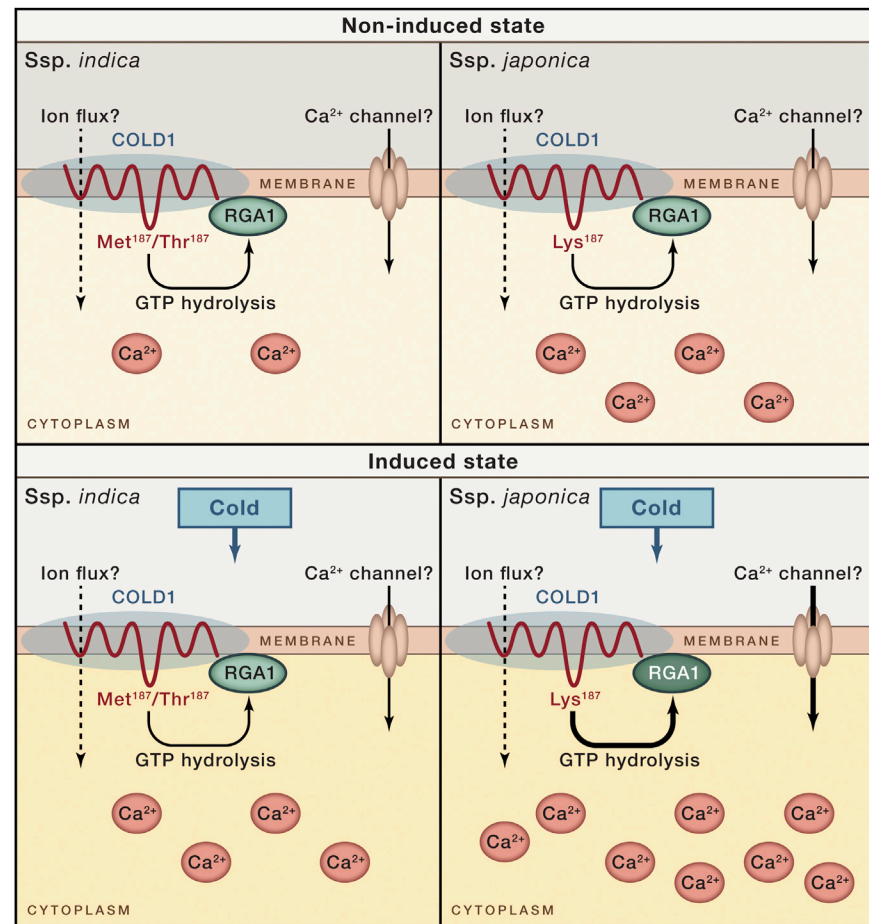
Rice (*Oryza sativa* L.) is one of the most important staple food crops consumed by half of the world's population. Rice is extensively cultivated on every continent in more than 100 countries (Juliano, 2003). Due to its diverse growing locations and climatic factors, rice is exposed to many biotic and abiotic stresses, which affect the physiological status, thereby affecting its overall metabolism (da Cruz et al., 2013). In particular, cold stress adversely affects the rice plants at their germination, vegetative growth, and reproductive stages, leading to severe yield reduction. Understanding the mechanisms of and improving the cold tolerance of rice is therefore of eminent importance for feeding the world's population. During a thousand years of rice domestication, two major genotypes have been bred and cultivated widely by the rice farmers: *japonica*, which exhibits superior cold tolerance and *indica*, with a higher yield.

In this issue of *Cell*, Ma et al. (2015) report the identification of a quantitative trait locus (QTL) named *COLD1* that confers cold tolerance in *japonica* rice. *COLD1* was identified in recombinant inbred lines generated from a cross between cold-tolerant Nipponbare (*japonica*) and cold-sensitive 93-11 (*indica*) cultivars. Ma et al. performed a fine mapping of *COLD1* by analyzing three near isogenic lines containing the *COLD1*<sup>NIP</sup> locus in the 93-11 background. This led to the identification of a single-nucleotide polymorphism, SNP2, originating from Chinese wild rice relative *Oryza rufipogon*. Quite excitingly, this one nucleotide change was responsible for conferring cold tolerance in *japonica* rice. The authors further corroborated this finding by genetic complementation and overexpression studies. This extensive study

provides fundamental insights of how cold tolerance was subjected to artificial selection during rice breeding.

*COLD1* appears to represent a nine transmembrane domain protein that is

related to *Arabidopsis* GTG1 (Pandey et al., 2009; Jaffé et al., 2012). In *Arabidopsis*, two genes (GTG1/GTG2) encode homologs of *COLD1* that have been implemented in ABA responses and plant



**Figure 1. Model of *COLD1* Function in Rice Cold Tolerance**

The two alleles of the membrane protein *COLD1* from the cold-sensitive *indica* and the more cold-tolerant *japonica* cultivars of rice differ in one amino acid of the third membrane-spanning domain (japLys187 versus indMet187/Thr187). This difference coincides with elevated cytoplasmic  $\text{Ca}^{2+}$  concentration in *japonica* compared to *indica*. Cold stress accelerates the GTPase activity of the G-protein  $\alpha$  subunit 1 (RGA1) upon RGA1 interaction with *COLD1*<sup>ind</sup>, but not with *COLD1*<sup>ind</sup>, further increasing cytoplasmic  $\text{Ca}^{2+}$  concentration in *japonica*. Details of the ion conductivity of *COLD1* and the identity of the  $\text{Ca}^{2+}$  channel(s) involved remain to be established.

development (Pandey et al., 2009; Fujisawa et al., 1999; Jaffé et al., 2012). An important question arising from this study is how COLD1 (and its plant homologs) may contribute to cold tolerance. Here, the topology of COLD1 suggests that it may function as an ion-conducting protein. The authors observe the localization of COLD1 in the endoplasmic reticulum and the plasma membrane. Interestingly, a recent study characterized a similar mammalian protein as being resident in the Golgi and functioning as a cellular Golgi pH regulator in Chinese hamster (*Cricetulus griseus*). This protein was found to be involved in Golgi acidification and functioning as a voltage-dependent anion channel (Maeda et al., 2008). Remarkably, Ma et al. reported an elevated basal  $\text{Ca}^{2+}$  concentration in rice plants expressing the cold-tolerant COLD1 allele. Moreover, they observed temperature-dependent changes in the protein structure of COLD1. These findings make it tempting to speculate that COLD1 might convey a specific physical parameter represented by temperature into changes in cellular  $\text{Ca}^{2+}$  concentrations. This  $\text{Ca}^{2+}$  signal would then trigger plant adaptation to the environmental cue accordingly (Figure 1).

The study by Ma et al. also provides evidence that COLD1 interacts with the

rice G protein  $\alpha$  subunit 1 (RGA1), suggesting that COLD1 might be involved in G-protein-dependent signal transduction. Importantly, they also demonstrated that COLD1<sup>jap</sup> from the cold-tolerant *japonica* cultivar, but not the allele from the cold-sensitive *indica* cultivar, accelerated the RGA1 GTPase activity. Similarly, truncated protein COLD1<sup>Δjap</sup> did not infer cold tolerance. In line with the potential contribution of COLD1<sup>jap</sup> to  $\text{Ca}^{2+}$  signaling, voltage-clamp recording in *Xenopus* oocytes revealed that COLD1 affected the influx of cations such as  $\text{Ca}^{2+}$  in the presence of RGA1. This observation suggests that the cold-stimulated inward current may originate from  $\text{Ca}^{2+}$ -dependent interaction between them.

These findings raise interesting questions considering the cross kingdom conservation of COLD1. It will be most interesting to address whether this protein may function as a temperature-regulated ion channel by analyzing COLD1 currents in reconstituted lipid bilayers. In addition, how do different alleles of COLD1 influence cytoplasmic  $\text{Ca}^{2+}$  concentrations? From this perspective, further elucidating the subcellular localization of COLD1 would be important for providing insights into its role in regulating cellular ion homeostasis. Another intriguing question is whether the elevated

$\text{Ca}^{2+}$  concentration in COLD1<sup>jap</sup> plants directly triggers enhanced cold tolerance. Would this mean that increasing resting  $\text{Ca}^{2+}$  levels and cold-induced  $\text{Ca}^{2+}$  release could be sufficient to cope with cold environment? Overall, this work may pave the way to tackle the food production insufficiency due to environmental changes and may contribute to food security by stabilizing the yield of a major crop that nurtures a large human population on this planet.

## REFERENCES

- da Cruz, R.P., Sperotto, R.A., Cargnelutti, D., Adamski, J.M., de FreitasTerra, T., and Fett, J.P. (2013). Food and Energy Security 2, 96–119.
- Fujisawa, Y., Kato, T., Ohki, S., Ishikawa, A., Kitano, H., Sasaki, T., Asahi, T., and Iwasaki, Y. (1999). Proc. Natl. Acad. Sci. USA 96, 7575–7580.
- Jaffé, F.W., Freschet, G.E., Valdes, B.M., Runions, J., Terry, M.J., and Williams, L.E. (2012). Plant Cell 24, 3649–3668.
- Juliano, B.O. (2003). Rice in Human Nutrition (Food and Agriculture Organization of the United Nations), pp. 1–15.
- Ma, Y., Dai, X., Xu, Y., Luo, W., Zheng, X., Zeng, D., Pan, Y., Lin, X., Liu, H., Zhang, D., et al. (2015). Cell 160, this issue, 1209–1221.
- Maeda, Y., Ide, T., Koike, M., Uchiyama, Y., and Kinoshita, T. (2008). Nat. Cell Biol. 10, 1135–1145.
- Pandey, S., Nelson, D.C., and Assmann, S.M. (2009). Cell 136, 136–148.

# Monkeys in a Prisoner's Dilemma

Ju Tian<sup>1,\*</sup> and Naoshige Uchida<sup>1,\*</sup>

<sup>1</sup>Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

\*Correspondence: jutian@fas.harvard.edu (J.T.), uchida@mcb.harvard.edu (N.U.)

<http://dx.doi.org/10.1016/j.cell.2015.02.049>

Haroush and Williams trained pairs of monkeys to play in a prisoner's dilemma game, a model of social interactions. Recording from the dorsal anterior cingulate cortex (dACC), they find neurons whose activity reflects the anticipation of the opponent's yet unknown choice, which may be important in guiding animals' performance in the game.

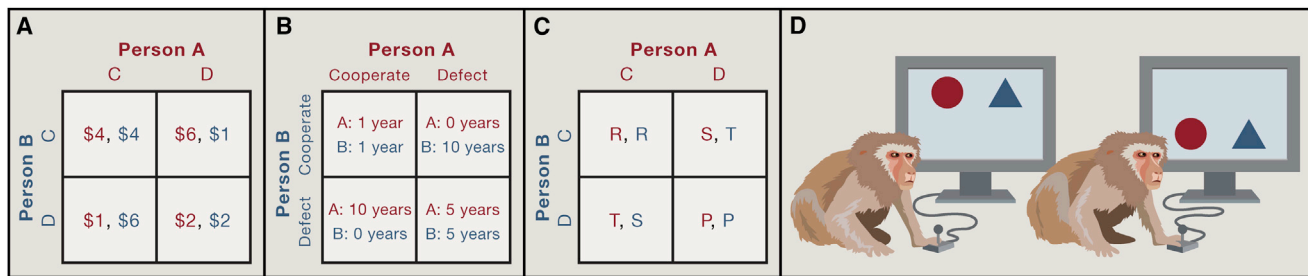
Imagine that you are playing the following game against a stranger. Each of you has to choose the option C or D without knowing which option your opponent will choose. Your outcome will depend both

on your own decision and your opponents, as outlined on a table (or a "payoff matrix"; Figure 1A). If both of you choose C, you both get \$4. If both choose D, both get \$2. However, if one chooses C and

the other D, the former gets the biggest reward (\$6) while the latter gets the smallest (\$1). Which option would you choose?

Here is one way to think. Assuming that your opponent chooses C, you get \$4





**Figure 1. Prisoner's Dilemma Game**

(A) An example payoff matrix of a game. There are two players (A and B), and each has to choose one of two options (C and D). Each makes a choice without knowing what the opponent will do. Depending on one's own and the other's choices, each receives an outcome defined by the payoff matrix (e.g., the number indicates dollar amount that each receives).

(B) The original prisoner's dilemma game. Each prisoner is asked to either defect the other by testifying that the other committed the crime or cooperate with the other by being silent.

(C) A payoff matrix in a general form.  $T > R > P > S$  and  $2R > T + S$  define a prisoner's dilemma. The latter criterion guarantees that the players cannot escape the dilemma simply by taking turns.

(D) Two monkeys played an iterated prisoner's dilemma game (defined by A in Haroush and Williams [2015]).

if you choose C and \$6 if you choose D. So you should choose D. Assuming that the opponent chooses D, you get \$1 if you choose C and \$2 if you choose D. So you should choose D again. The answer is simple! No matter what the opponent will do, you are always better off choosing D.

A closer look might make you unhappy though. Choosing D actually results in the worst outcome in terms of the total gain ( $2+2 < 1+6 < 4+4$ ). Moreover, both players choosing C (\$4) is better than both players choosing D (\$2). Why not both choose C? Well, if your opponent knows that you will choose C, he or she might betray or defect you (i.e., choose D) to get a larger reward! "Cooperation" is needed for the common good.

This type of game is called a prisoner's dilemma (PD), which was named after the famous example of prisoners negotiating with attorneys (Figure 1B) (Camerer, 2003). To be a PD game, the payoff matrix has to fulfill specific criteria (Figure 1C). It is the mathematical structure of the payoff matrix that generates the sense of cooperation and defection. In other words, one need not be told that C is cooperation and D is defection.

Game theory studies what happens when people—or genes or nations—interact (Camerer, 2003; Morgenstern and Von Neumann, 1953). It provides the strategy that a self-interested "rational" agent must follow in such situations. In the case of the game described above, game theory predicts that both players will choose D (mutual defect) since there

is no incentive for each player to move away from it (that is, mutual defection is the only "Nash equilibrium" in the PD game). Contrary to this reasoning, when humans play the PD game, about half of the players cooperate (Camerer, 2003). When the games are repeated with the same stranger (iterated PD), cooperation starts high and then decreases over time (Camerer, 2003; Rilling et al., 2002). When non-human animals play PD games, fewer but some cooperative behaviors have been observed (Stevens and Hauser, 2004). The PD game has been regarded as the *E.coli* of social psychology (Axelrod, 1997): it mimics many real-world dilemmas and is thought to be a good model to study the emergence and development of cooperative behavior. Yet, very little is known about the neural underpinnings of PD games (Behrens et al., 2009; Fehr and Camerer, 2007; Rilling et al., 2002). To address this question at a single-neuron level, in this issue of *Cell*, Haroush and Williams (2015) trained monkeys, not humans, to play in an iterated PD game (Figure 1D) (Haroush and Williams, 2015).

In their study, the monkeys sit side by side and make decisions sequentially to obtain different amounts of juice instead of money. They cannot see the other's choice until both have made their selections. Contrary to the game theoretic prediction, the monkeys choose C ("cooperation") in 34.7% of trials. Note that choosing C does not necessarily mean that the monkeys understand the concept of "cooperation" or even

aim for mutual benefits; in this task, it is hard to know whether the monkeys know the amount of juice the opponent got. Note also that the monkeys have to learn the payoff matrix by playing (that is, no explicit explanation of the payoff matrix could be given). Nevertheless, the monkeys choose C more often if the other chooses C in the preceding trial and less so if the other chooses D, similar to how humans perform in this game. Furthermore, when a monkey plays either with a computer or with a monkey partner in a separate room, the overall probability of choosing C greatly decreases, suggesting that social contexts affect their choices. Lastly, to probe whether the monkeys have good understanding of the payoff matrix, in some trials, the monkey is informed of the opponent's choice before it makes a decision. In these trials, the monkey chooses D more than 90% of the time when the opponent had already chosen D.

The authors then recorded the activity of single neurons in the dorsal anterior cingulate cortex (dACC). The ACC is subdivided into the dorsal and ventral parts (dACC and vACC, respectively); dACC is thought to be involved in reward-guided decisions and processing cognitive conflicts, whereas vACC is involved in social emotions and social interests (Behrens et al., 2009; Rilling et al., 2002; Rudebeck et al., 2006; Somerville et al., 2006). They find two non-overlapping neuronal populations whose activity co-fluctuates with either the monkey's own choice or the

opponents' yet-unknown choice. Specifically, 27.6% of the recorded neurons encode the opponent's choice (versus 11.4% for self-choice) during the post-selection period and 7% (versus 15.7% for self-choice) during the pre-selection period. Note that both of these periods are before the opponent's decision is revealed to the monkey, suggesting that these activities are related to prediction or anticipation of the opponent's choice. Based on the activity of a population of other-predicting neurons, it is possible to "decode" the opponent's choice with high precision (79.4%). Importantly, the number of other-predicting neurons decreases when two monkeys play in separate rooms.

Further analyses help to exclude the possibility that other-predicting neurons are encoding other task features. For example, based on the payoff matrix, the monkey receives an overall larger reward (four or six drops of juice) when the opponent chooses C compared to when the opponent chooses D (one or two drops). Could these "other-predicting" neurons in fact encode expected self-reward? Their results suggest that this is not the case.

They further show that disrupting the dACC activity by applying a strong electrical current during the pre-selection period decreases the odds of choosing C. This effect is most prominent in trials when the opponent chose C in the previous trial. It is unclear, however, whether this behavioral effect is due to the alteration of other-predicting neurons; most other-predicting neurons are active after rather than before selection. Instead, other-predicting neurons may contribute to learning for future trials. Further efforts are required to elucidate how other-predicting neurons contribute to choices and what aspects of social interactions or prior experience drive their activity. Furthermore, how electrical stimulation of the dACC, which may perturb the activity of other interconnected areas, leads to less "cooperative" choices remains to be further investigated. Finally, what really makes the difference between two monkeys sitting side by side versus playing in separate rooms? This last question may provide insights into what defines "social."

Haroush and Williams (2015) provide a powerful experimental system to study the neural mechanisms underlying

social decision making. The abilities to record single-neuron activities and to manipulate their activities offer unprecedented opportunities to unravel intricate brain processes underlying aspects of social interactions.

## REFERENCES

- Axelrod, R. (1997). *The Complexity of Cooperation* (Princeton University Press).
- Behrens, T.E.J., Hunt, L.T., and Rushworth, M.F.S. (2009). *Science* 324, 1160–1164.
- Camerer, C.F. (2003). *Behavioral Game Theory* (Princeton University Press).
- Fehr, E., and Camerer, C.F. (2007). *Trends Cogn. Sci.* 11, 419–427.
- Haroush, K., and Williams, Z.M. (2015). *Cell* 160, this issue, 1233–1245.
- Morgenstern, O., and Von Neumann, J. (1953). *Theory of games and economic behavior* (Princeton University Press).
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). *Neuron* 35, 395–405.
- Rudebeck, P.H., Buckley, M.J., Walton, M.E., and Rushworth, M.F.S. (2006). *Science* 313, 1310–1312.
- Somerville, L.H., Heatherton, T.F., and Kelley, W.M. (2006). *Nat. Neurosci.* 9, 1007–1008.
- Stevens, J.R., and Hauser, M.D. (2004). *Trends Cogn. Sci.* 8, 60–65.

# The Role of Chromosome Domains in Shaping the Functional Genome

Tom Sexton<sup>1,\*</sup> and Giacomo Cavalli<sup>2,\*</sup>

<sup>1</sup>Institute of Genetics and Molecular and Cellular Biology (IGBMC), 1 rue Laurent Fries, 67404 Illkirch, France

<sup>2</sup>Institute of Human Genetics (IGH), 141 rue de la Cardonille, 34396 Montpellier, France

\*Correspondence: [sexton@igbmc.fr](mailto:sexton@igbmc.fr) (T.S.), [giacomo.cavalli@igh.cnrs.fr](mailto:giacomo.cavalli@igh.cnrs.fr) (G.C.)

<http://dx.doi.org/10.1016/j.cell.2015.02.040>

The genome must be highly compacted to fit within eukaryotic nuclei but must be accessible to the transcriptional machinery to allow appropriate expression of genes in different cell types and throughout developmental pathways. A growing body of work has shown that the genome, analogously to proteins, forms an ordered, hierarchical structure that closely correlates and may even be causally linked with regulation of functions such as transcription. This review describes our current understanding of how these functional genomic “secondary and tertiary structures” form a blueprint for global nuclear architecture and the potential they hold for understanding and manipulating genomic regulation.

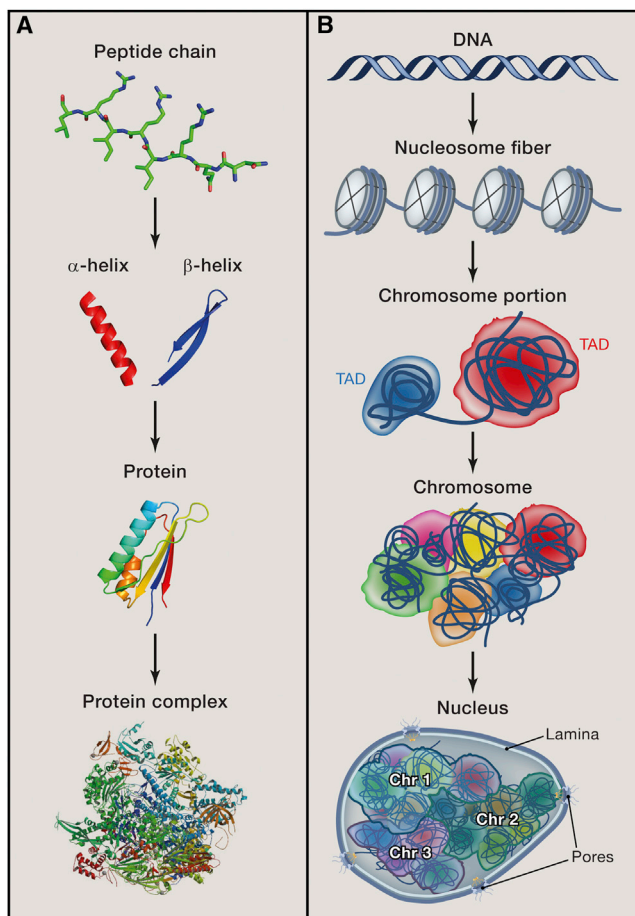
Eukaryotic genomes must be tightly folded and packaged to be contained within cell nuclei. Since initial observations of heterochromatin by Emil Heitz in the 1930s, it has become more and more appreciated that this packaging is highly organized and may be closely linked to transcriptional control. Over the last two decades, many studies have assessed the spatial proximity and nuclear organization of specific genomic loci, using microscopic techniques, such as fluorescent in situ hybridization (FISH), or molecular biology techniques, such as chromosome conformation capture (3C). Collectively, these studies demonstrated a correlation between chromatin topology and underlying gene activity, without resolving whether chromosome folding is a cause or consequence of genomic functions (Cavalli and Misteli, 2013; de Laat and Duboule, 2013).

Topology and activity appear linked at different scales within the nucleus. At the kilobase-to-megabase scale, distal regulatory elements such as enhancers were found to come into direct contact with their target genes via chromatin loops (Palstra et al., 2003). At the megabase scale, genes were observed to significantly co-occupy functional sites within the nucleus, such as foci of Polycomb proteins (Bantignies et al., 2011) or of active RNA polymerase (Schoenfelder et al., 2010), specifically in cells where the genes have the same activity. At the scale of the whole nucleus, chromosomes occupy discrete territories, which are non-randomly organized to place gene-poor chromosomes in the predominantly heterochromatic periphery and gene-rich regions in the euchromatic interior. The transcriptional activity of specific genes has been correlated with their nuclear positioning relative to the periphery, and more specifically the repressive nuclear lamina (Peric-Hupkes et al., 2010), as well as to their position relative to the bulk of the chromosome territory (Chau-meil et al., 2006). Intriguing recent work has even decoupled chromatin decondensation from transcriptional activation, showing that opening chromatin without concomitant gene activation is sufficient for relocalization of genes to the nuclear

interior (Therizols et al., 2014). Overall, these case studies support a hierarchical, multi-scale model where expression of a gene may influence or be influenced by its local chromatin interactions, its associations with other potentially coordinately controlled genes and the regulatory environment provided by its nuclear location.

Average conformations of chromatin have been more systematically characterized by coupling 3C to high-throughput sequencing (Hi-C) to derive large catalogs of pairwise chromatin interactions within populations of nuclei (Lieberman-Aiden et al., 2009). Initial, lower-resolution Hi-C studies demonstrated that active chromatin predominantly associates with other active regions, and repressed chromatin associates with other silent regions with little inter-mixing of the two types (Lieberman-Aiden et al., 2009). More recently, high-resolution chromatin interaction maps revealed that metazoan genomes fold into distinct modules called physical domains or topologically associated domains (TADs), whereby genomic interactions are strong within a domain but are sharply depleted on crossing the boundary between two TADs (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). The presence of TADs is less clear for non-animal species. Although Hi-C is unable to give any information on TAD dynamics or cell-to-cell variability, the domains identified correlate well with many markers of chromatin activity, such as histone modifications and replication timing (Dixon et al., 2012; Sexton et al., 2012). TADs can also contain coordinately regulated genes (Le Dily et al., 2014; Nora et al., 2012). The described organization of the genome into functional domains containing different types of chromatin (Ernst et al., 2011; Ho et al., 2014) thus reflects the average folded state of the chromosome.

TADs appear to form the modular basis for higher-order chromosomal structures (Sexton et al., 2012), which in themselves may be built up from key stabilizing interactions between regulatory elements (Giorgetti et al., 2014). Such an arrangement is reminiscent of protein folding, whereby hierarchical stabilization



**Figure 1. Analogous Hierarchical Organization of Protein and Genome Structure**

(A and B) Primary structures comprising the amino acid or nucleotide sequence (packaged into a nucleosomal fiber in eukaryotic chromatin) on a single polymeric chain form locally stabilized interactions to fold into secondary structures, such as polypeptide alpha-helices or beta-sheets, or chromatin TADs. These domains in turn hierarchically co-associate to form a tertiary structure of a protein or chromosome. The co-associations of multiple, separately encoded subunits forms the final quaternary structure of a protein complex or entire genome. Protein structures taken or derived from the RCSB database (PDB 2KVQ, or 4BBR for quaternary structure).

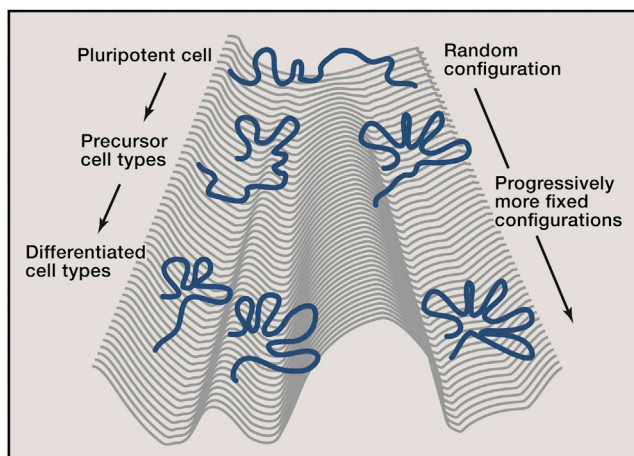
of secondary structures such as alpha-helices leads to the final tertiary structure, whose conformation is crucial to protein function (Figure 1). Genome folding is not as rigidly or thermodynamically defined as protein structure—single-cell experiments reveal a high variability of adopted genomic configurations (Nagano et al., 2013; Noordermeer et al., 2011a). Further, it has not been shown that a specific chromosome structure is essential for genomic functions. However, considering chromosome topology as a principle of folding, and TADs as chromosomal secondary structures, is a useful starting analogy. Here, we discuss the relationship between DNA sequence (primary structure), genomic sub-structures such as TADs (secondary structure), overall chromosome folding (tertiary structure), and genome function, positing that TADs and other localized structures form a blueprint for coordinated genome control.

### Chromatin Loops in Gene Regulation

Seminal studies of the beta-globin locus showed that the globin gene promoter more frequently interacted with distal enhancers than intervening sequence, specifically in erythroid tissue where the gene was transcribed (Palstra et al., 2003). Such results were confirmed for other enhancer-promoter combinations (Kieffer-Kwon et al., 2013; Li et al., 2012; Sanyal et al., 2012) and suggest that chromatin looping brings genes and their regulatory elements in close proximity. For simplicity, we will also refer to these phenomena as loops, although in many cases they are more likely to represent a statistical ensemble of transient contacts than true stable structures (Giorgetti et al., 2014). Many enhancer-promoter combinations share binding of common transcription factors, and enhancers are also frequently transcribed, especially when involved in interactions with target genes (Sanyal et al., 2012). Such chromatin loops are thus proposed to set up an “active chromatin hub,” providing a chromatin environment more permissive to transcription than factors bound directly to the promoter alone (Mousavi et al., 2013; Palstra et al., 2003). In support of this model, enhancer-promoter interactions within the human *OCT4* locus, a gene encoding a key pluripotency transcription factor, distinguish induced pluripotent stem cells from non-reprogrammed cells (Zhang et al., 2013). The non-reprogrammed cells had equivalent binding of the inducing factors at the promoter and enhancer but no *OCT4* expression. However, it remains an open question whether chromatin looping is a cause or consequence of transcriptional activation. Recent elegant experiments have engineered chromatin loops within the mouse beta-globin locus by exogenously targeting the dimerization domain of the transcription factor Ldb1, which is naturally present at the enhancers of the globin locus control region (Deng et al., 2012; Deng et al., 2014). These induced chromatin loops could partially rescue adult beta-globin expression in mutants for erythroid transcription factors (Deng et al., 2012) or stimulate fetal globin expression out of its normal developmental context (Deng et al., 2014). Chromatin topology can thus be causally linked to transcriptional regulation. As the globin genes are very highly expressed in erythroid tissues, it will be interesting to see the functional consequences of induced chromatin loops in less transcriptionally permissive genomic and cell-type contexts.

The beta-globin active chromatin hub is progressively formed during hematopoiesis (Palstra et al., 2003) and involves binding sites for erythroid-specific transcription factors (Drissen et al., 2004 for example), so enhancer-promoter contacts were proposed to occur exclusively in cells where the target gene is being transcribed. Although many cell-type-specific chromatin loops have been characterized from more systematic approaches (Heidari et al., 2014; Sanyal et al., 2012), evidence is also emerging that chromatin topology and transcriptional regulation can be temporally uncoupled. A recent analysis of the interaction profiles of a hundred *Drosophila* mesodermal enhancers found that more than 90% of the interactions were detectable before mesoderm specification and were commonly linked to genes with paused RNA polymerase (Ghavi-Helm et al., 2014). This result suggests that chromatin loops may commonly poise a gene for expression but that another signal is required for complete transcriptional firing. In support of this model, induced looping within the beta-globin locus rescued transcription





**Figure 2. Waddington Landscape of Chromatin Loop Configurations throughout Development**

Pluripotent cells able to form any lineage (top) have largely unstructured local chromatin topologies. Progressive lineage restriction throughout development, tracing paths through the landscape from top to bottom, may be accompanied by progressive constraint of the specific chromatin loop topologies as only a limited repertoire of enhancer-promoter contacts are permitted and fixed.

initiation, but not efficient elongation when the essential transcription factor GATA-1 was lacking (Deng et al., 2012). Furthermore, Hi-C analysis of a human fibroblast cell line showed conservation of enhancer-promoter interactions around responsive genes before and after treatment with the cytokine TNF- $\alpha$  (Jin et al., 2013).

These seemingly opposing views of enhancer-promoter chromatin loop dynamics may be reconciled by a Waddington landscape model of chromatin architecture (Figure 2). Non-expressed genes form more promiscuous contacts in pluripotent cells than in differentiated cells (de Wit et al., 2013; Splinter et al., 2011). Repertoires of tissue-specific interactions may then be set up in precursor cells as their differentiation potential is restricted, effectively limiting the sets of genes with a permissive chromatin environment for further induction. Fully differentiated cells may then benefit from their pre-formed active chromatin hubs for rapid transcriptional responses to appropriate signals. Although this model has yet to be formally assessed, chromatin states themselves exhibit a similar progressive developmental restriction (Zhu et al., 2013). Furthermore, there is more tissue-type variation in the chromatin states of enhancers than of promoters (Ernst et al., 2011). Finally, a recent analysis has suggested that enhancer-promoter interactions are variable in different cell types (He et al., 2014). Together, these data suggest that enhancers carry a large regulatory potential, and although the mechanistic details of *when* and *how* they stimulate transcription are not yet clarified, chromatin loops appear a ubiquitous means of relaying enhancer-promoter communication.

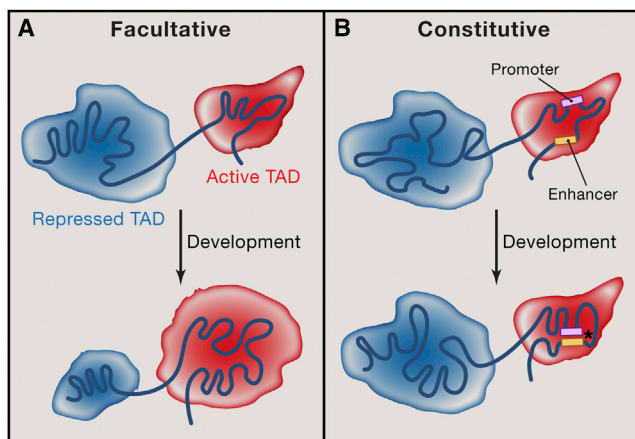
### Architectural Chromatin Loops—Building up the Secondary Structures

In addition to specific transcription factors, ubiquitous proteins have also been linked to chromatin loops, in particular the insu-

lator protein CTCF (Splinter et al., 2006), the cohesin complex (Hadjur et al., 2009), and the general co-activating Mediator complex (Kagey et al., 2010). Mediator is predominantly found at loops between promoters and enhancers and between promoters, in agreement with its general activation role (Conaway and Conaway, 2011). Consistently, Mediator-linked interactions are more cell-type-specific (Phillips-Cremins et al., 2013). In contrast, CTCF tends not to be present at enhancer-promoter loops. It is more commonly associated with constitutive, longer-range chromatin interactions (Phillips-Cremins et al., 2013; Sanyal et al., 2012), although some cell-type-specific CTCF-mediated interactions have been reported (Hou et al., 2010). CTCF is enriched at TAD borders (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012), and CTCF-mediated loops are implicated in maintenance of TAD structure (Giorgetti et al., 2014) and are thus believed to play a more fundamental architectural role in chromosome folding. Various case studies have implicated CTCF-mediated loops in insulator function, preventing communication between distal regulatory elements (Kurukuti et al., 2006 for example). However, many CTCF sites have recently been shown not to be a barrier to enhancer-promoter interactions (Sanyal et al., 2012). The functional consequences of these more developmentally stable chromatin architectures are thus likely to be complex and context-dependent. Similarly, CTCF binding alone cannot account for TAD border function (discussed in more detail in later sections). Cohesins are associated with both cell-type-specific enhancer-based loops and constitutive, CTCF-mediated loops, although both types of loops can also be cohesin-independent (DeMare et al., 2013; Phillips-Cremins et al., 2013). In agreement, cohesin has been shown to interact with CTCF (Rubio et al., 2008) and forms direct complexes with Mediator (Kagey et al., 2010) and certain transcription factors (Wei et al., 2013). The cohesin complex comprises a ring structure that physically maintains sister chromatid attachment after DNA replication (Nasmyth and Haering, 2009). Though yet to be demonstrated, a similar structure could be envisioned to stabilize chromatin loops on cohesin recruitment. Abrogation of cohesin causes perturbation of chromatin loops with subsequent effects on transcriptional control (Hadjur et al., 2009; Seitan et al., 2013; Sofueva et al., 2013; Zuin et al., 2014). Overall, chromatin loops appear important for the possibly inter-linked functions of transcriptional regulation and maintenance of higher-order chromosome folding. A full proteomic appraisal of the factors present at chromatin loops may help us better understand how they are recruited to their specific sites in a developmental context and how and when they are able to effect looping.

### Chromosomal Secondary Structures—“Facultative” and “Constitutive” TADs

The three-dimensional organization of many metazoan genomes into discretely folded kilobase-to-megabase sized TADs is particularly striking due to their agreement with many linear (or one-dimensional) measurements of chromatin activity; for example, histone modifications (Dixon et al., 2012; Sexton et al., 2012), coordinated gene expression (Le Dily et al., 2014; Nora et al., 2012), lamina association (Dixon et al., 2012), and DNA replication timing (Dixon et al., 2012; Pope et al., 2014).



**Figure 3. Facultative and Constitutive TAD Models of Regulated Developmental Gene Expression Programs**

(A) Active (red) and repressed (blue) chromatin domains form separate facultative TADs which spatially segregate their regulatory environments. During development, some genes are activated and leave the repressive TAD to enter the growing facultative active TAD by a shift in the boundary between TADs. (B) Boundary positions do not change in constitutive TADs. Gene expression changes are effected via altered intra-TAD chromatin interactions; for example by developmental stage-specific presence of enhancer-promoter chromatin loops (asterisk; positions of sequences participating in this loop in both cell types are highlighted in yellow and pink).

TADs thus appear to be chromosomal secondary structures that reflect a tendency to divide the genome into distinct, autonomously regulated regions. This model is supported by the finding that TADs determine the scope of most enhancers' activities (Ghavi-Helm et al., 2014; Shen et al., 2012; Symmons et al., 2014). The mechanisms of TAD establishment and maintenance are largely unknown. In particular, a critical issue to be resolved is whether TADs constitute a structural blueprint that defines chromosome architecture within which gene regulatory changes are overlaid, or are themselves dynamically built by transcriptional silencing or activation machineries. A case in point for TAD organization by transcription arises from studies aimed at understanding the spatial and temporal collinearity of mouse Hox gene expression. These genes are sequentially activated during development, and according to anterior-posterior body position, in order along the chromosomal fiber. The active genes are marked by trimethylation of lysine-4 of histone H3 (H3K4me3) and the silent regions are coated with trimethylation of lysine-27 of histone H3 (H3K27me3). Hox gene activation is accompanied by a transition in the chromatin modification (Soshnikova and Duboule, 2009). Strikingly, the Hox gene loci form distinct topological domains which mirror these chromatin domains precisely, with the active domain expanding and the silent domain shrinking according to collinear gene activation (Noordermeer et al., 2011b). Such a dynamic model of chromosome topology implies that "facultative TADs" spatially confine co-regulated genomic regions but may actually be defined by the underlying transcriptional activity and/or chromatin state (Figure 3A). However, ablation of H3K27me3 in mouse ES cells by knockout of the Polycomb group gene *Eed* had no effect on TAD structures around the X-inactivation locus (Nora et al., 2012). Further, genome-wide comparisons of TADs in disparate

mouse and human cell lines and tissues revealed that most TADs seem invariant with cell type (Dixon et al., 2012). Although many TADs at gene deserts or clusters of ubiquitously expressed housekeeping genes would not necessarily be expected to change in these different cell types, the large number of "constitutive TADs" suggests that many are genuine chromosomal secondary structures. These may thus represent a ground state spatial configuration on which subsequent regulatory features are overlaid (Figure 3B). In support of this view, entire TADs containing coordinately responsive genes to progesterone treatment can be structurally re-modeled while their borders remain unchanged (Le Dily et al., 2014). In between these extreme views of chromosome topology, high-resolution analysis of a handful of TADs during ES cell differentiation identified them to be predominantly stable but noted developmental dynamics of smaller "sub-TADs" within them (Phillips-Cremins et al., 2013). As the resolution of genome-wide chromatin interaction maps improves, so will our appreciation of the interplay between developmentally stable and dynamic chromosomal secondary structures and of the cause-effect relationships between TADs and genome function.

### Establishing, Maintaining, and Re-Building Chromosomal Secondary Structures

Despite (or perhaps because of) their many correlations with different epigenomic features, unravelling the causal factors in TAD establishment and maintenance remains a challenge. TAD borders in *Drosophila* are very significantly associated with binding of various insulator proteins (Hou et al., 2012; Sexton et al., 2012); CTCF is the only one of these factors conserved in mammals and is also enriched at constitutive TAD borders (Dixon et al., 2012). However, the full link between insulators and chromosome topology remains unclear—in one genome-wide study around a quarter of TAD borders did not contain CTCF and only 15% of CTCF binding sites were present at TAD boundaries (Dixon et al., 2012). Further, knockdown of CTCF in a human cell line caused an increase in the chromatin interactions spanning TAD borders but did not completely disrupt TAD organization (Zuin et al., 2014). This result is consistent with the persistent demarcation of H3K27me3 domains in *Drosophila* on CTCF knockdown (Van Bortle et al., 2012). In mammals, but not *Drosophila*, cohesin is also significantly found at TAD borders, although again the majority of binding sites are not at borders (Nora et al., 2012; Phillips-Cremins et al., 2013). Furthermore, cohesin abrogation in post-mitotic cells has no (Seitan et al., 2013; Zuin et al., 2014) or weak (Sofueva et al., 2013) effects on TAD border function. Although the effects of persisting levels of functional CTCF or cohesin cannot be ruled out in these studies, collectively it appears that these so-called "architectural proteins" contribute to the functional organization of the genome but that chromosomal secondary structures are largely epistatic to them.

TAD borders are also highly enriched in transcriptionally active genes (Dixon et al., 2012; Hou et al., 2012; Sexton et al., 2012), although the presence of borders at silent domains and the majority of transcribed genes residing inside domains mean that transcription alone cannot account for TAD organization. However, the known effects of RNA polymerase binding and

elongation on local DNA topology (Lavelle, 2014) suggest that gene expression programs and chromatin organization could have a profound effect on higher-order chromosome folding. In active chromatin, not only do enhancers contact promoters, but the promoters of expressed genes also contact each other (Li et al., 2012; Sanyal et al., 2012), and these interactions could favor TAD formation. Furthermore, active yeast genes form loops between their start and end sites to coordinate initiation and termination events, and this phenomenon appears to be conserved for at least some mammalian genes (Grzechnik et al., 2014). Transcription units could conceivably form a type of facultative mini-TAD. In support of this, active topological domains are smaller and more structurally complex than silent domains (Hou et al., 2012; Sexton et al., 2012; Sofueva et al., 2013). TAD borders are also enriched in housekeeping genes (Dixon et al., 2012). Evidence is mounting that housekeeping or widely expressed genes have fundamentally different regulatory sequences and chromatin states than developmentally regulated genes (Rach et al., 2011; Schauer et al., 2013; Zabidi et al., 2014). It will be interesting to see if these features, rather than maintained transcription per se, could contribute to TAD organization.

The tendency of chromatin domains of the same type to establish strong interactions is not limited to active chromatin domains. Polycomb domains are formed by clusters of Polycomb-bound sites that form preferential interactions, both intra-TAD (Lanzuolo et al., 2007; Schuettengruber et al., 2014) and inter-TAD (Bantignies et al., 2011; Sexton et al., 2012). Likewise, HP1-bound heterochromatin is involved in specific interactions (Csink and Henikoff, 1996; Sexton et al., 2012). Recent polymer physics-based modeling showed that the simple assumption of the existence of homotypic interactions between domains formed of these chromatin types is sufficient to generate polymer structures mimicking those shown in Hi-C contact maps (Jost et al., 2014). This result suggests that chromatin components of each type of chromatin domain may contribute to establish TADs. The role of boundary factors such as CTCF could thus be to strengthen the stability of the boundaries between domains of different chromatin types or to sharpen their localization.

One experimental test that has appreciably disrupted topological domain structure was the deletion of a 58-kb region spanning a TAD border within the X-inactivation locus. This perturbation resulted in complete loss of border function and the establishment of a new TAD border approximately 50 kb downstream of the deletion site (Nora et al., 2012). Interestingly, the de novo creation of a TAD boundary near to the deleted one was predicted from physical models and suggests that the chromosomes of many genomes have an intrinsic tendency to fold into topological domains (Giorgetti et al., 2014). Thus, at least some topological domain boundaries have a genetic component. Although it has yet to be demonstrated experimentally, disease phenotype association studies have also suggested that around one tenth of human pathologies caused by genomic deletions could involve perturbed topological domain function (Ibn-Salem et al., 2014). Finer dissection of the *cis*-sequence requirements of TAD borders and testing their function outside of their usual genomic contexts, should be fruitful in explaining

the mechanistic basis of chromosome organization and in enabling chromosome domain engineering.

Global chromosome structure is regulated throughout the cell cycle. Hi-C experiments have further shown that, whereas TAD organization is largely conserved throughout interphase, the domains are lost during mitosis (Naumova et al., 2013). The robust detection of conserved TADs in early G1 cells suggests that they can be efficiently re-built. Characterization of the proteins and chromatin marks that persist on mitotic chromosomes, the so-called “bookmarking” factors, is an area of current intense study, which may yield some clues as to how TADs can be established at each cell cycle (Zaret, 2014). For example, it has been shown in *Drosophila* that the Polycomb group protein PSC persists on only a subset of binding sites during mitosis and that these are predominantly interphase TAD boundaries (Follmer et al., 2012). However, it is unclear how this bookmarking is regulated, if or how it controls TAD organization, or how the many TADs that are not mitotically bound by PSC are regulated. DNA damage and the chromatin remodeling accompanying its repair are also likely to affect the organization of the associated TADs. Although previous results have shown that heterochromatin domains have different induced mobility and/or repair mechanisms in response to double-stranded breaks (Chiolo et al., 2011; Lemaître et al., 2014), it is still unknown how TADs are maintained or restored in different nuclear environments. Overall, genetic elements, transcription, and the binding of architectural proteins have all been correlated with TAD borders. Future research should tease out whether they are causes or consequences of TAD folding, how these factors interplay in such organization, and their roles in re-building TADs after mitosis.

### Chromosomal Secondary Structures in Genome Evolution

TAD organization appears to be a conserved, but not universal phenomenon (Table 1); TADs are readily observed in *Drosophila* (Hou et al., 2012; Sexton et al., 2012) and mammalian (Dixon et al., 2012; Nora et al., 2012) genomes but are less clearly defined in *Arabidopsis* (Feng et al., 2014; Grob et al., 2014), *Plasmodium falciparum* (Ay et al., 2014), and yeasts (Duan et al., 2010; Tanizawa et al., 2010). Although more systematic chromatin interaction maps of different organisms are required to make further conclusions, it is interesting that species with clear TAD genomic organization match those with conservation of the insulator protein CTCF (Heger et al., 2012), further supporting its role as a genomic architectural protein. However, closer analysis of chromatin interaction maps of non-metazoan species reveals some topological domain-like organizations, such as the very large “structural domains” in *Arabidopsis* (Grob et al., 2014), or the tens of kilobase-sized “globules” in *Schizosaccharomyces pombe*, which correlate with the organization of convergent genes and cohesin binding sites (Mizuguchi et al., 2014). More strikingly, the chromosome of the bacterium *Caulobacter crescentus* also adopts TAD-like domains, which are highly sensitive to transcriptional activity and negative supercoiling (Le et al., 2013). Thus, genomic folding into potentially self-organized modules appears to be a common strategy for very diverse types of chromatin, perhaps reflecting an intrinsic ability for chromatin

**Table 1. Overview of the Absence or Presence of Chromosome Topological Domains, as Well as Their Observed Sizes, Based on Current Studies**

Organism	Evidence for TADs or Similar Domains	Domain Size	Methods Used	References
<i>C. crescentus</i>	Yes	30–420 kb	Hi-C and a sub-genome-wide derivative (5C)	<a href="#">Le et al., 2013</a>
<i>S. cerevisiae</i>	No	NA	A genome-wide 4C derivative	<a href="#">Duan et al., 2010</a>
<i>S. pombe</i>	Yes	50–100 kb	Hi-C	<a href="#">Mizuguchi et al., 2014</a>
<i>P. falciparum</i>	Only around a specific group of genes	10–50 kb	Hi-C	<a href="#">Ay et al., 2014</a>
<i>A. thaliana</i>	Controversial	> 1 Mb in one study; no TADs in another	Hi-C	<a href="#">Feng et al., 2014</a> ; <a href="#">Grob et al., 2014</a>
<i>D. melanogaster</i>	Yes	10–980 kb	Hi-C	<a href="#">Sexton et al., 2012</a>
<i>M. musculus</i>	Yes	100 kb–5 Mb	Hi-C, 5C	<a href="#">Dixon et al., 2012</a> ; <a href="#">Nora et al., 2012</a>
<i>H. sapiens</i>	Yes	100 kb–5 Mb in one study, 40 kb–3 Mb in another	Hi-C	<a href="#">Dixon et al., 2012</a> ; <a href="#">Rao et al., 2014</a>

to be compacted in a way that can be easily opened and re-condensed without entangling of chromosome fibers ([Lieberman-Aiden et al., 2009](#)). Until very recently, the TAD size of an organism appeared to scale with the average gene or chromosome length ([Table 1](#)). However, Hi-C coupled to extremely deep sequencing has identified human domains at a similar scale to that observed in *Drosophila* ([Rao et al., 2014](#)). Caution with respect to the resolutions afforded by different studies is thus required when trying to make cross-species comparisons of chromosome folding.

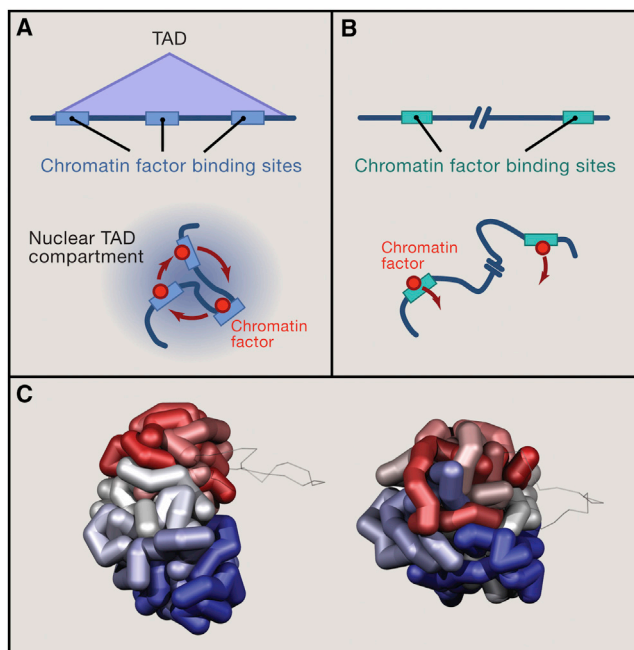
Comparison of mouse and human chromatin interaction maps revealed a high degree of TAD organization conservation around syntenic regions ([Dixon et al., 2012](#)). If these domains truly represent autonomously functional units of the genome, then rearrangements of whole TADs may be favored over ones that split TADs apart. Although such selection has not been formally proven, random P element insertions are highly enriched at TAD boundaries ([Hou et al., 2012](#)), suggesting that they may be genetic loci particularly susceptible or permissive to rearrangement events. It is also curious that distal human sequences which are syntenic in the mouse genome retain long-range chromatin interactions, tens of millions of years after the synteny break ([Véron et al., 2011](#)). This is not an isolated observation as Polycomb-dependent long-range contacts between Hox loci are conserved among fly species that diverged around 40 million years ago ([Bantignies et al., 2011](#)). Genome evolution could thus potentially be driven by re-arranging their secondary structures, analogous to the evolution of proteins by shuffling domain-coding exons ([Liu and Grigoriev, 2004](#)). Conversely, the spatial organization of TADs may also influence the sequence divergence within them. A recent comparative genomics study in *Drosophilid* species found that the dual transcription factor/Polycomb recruiter protein PHO bound only to consensus motif sequences outside of a Polycomb context but was able to bind far weaker motifs within TADs marked by H3K27me3 (a hallmark of Polycomb-mediated repression) ([Schuettengruber et al., 2014](#)). Of note, these Polycomb-linked PHO sites participated in stronger chromatin interactions, consistent with known looped interactions

between Polycomb group response elements ([Lanzuolo et al., 2007](#)). Such co-operative interactions within specific TADs were proposed to stabilize PHO binding, allowing a greater tolerance of motif sequence divergence ([Schuettengruber et al., 2014](#)). Thus DNA sequence appears to influence chromosome folding, and 3D chromosome structure in turn may influence sequence evolution ([Figures 4A and 4B](#)). These data call for more work in order to understand whether this principle may apply to the binding of a wide variety of transcription factors in eukaryotes.

### Toward Tertiary Chromosomal Structures

At current sequencing depths, Hi-C experiments are able to give fairly detailed views of TAD organization, but the resolution of longer-range (and interchromosomal) contacts is more limited. Although there is evidence to suggest that TADs hierarchically co-associate to build up larger chromosomal structures ([Sexton et al., 2012](#)), the precise nature of such spatial configurations remains mysterious. FISH studies of long-range gene co-associations in mouse erythroid cells or *Drosophila* embryos detected specific long-range interactions in only a few percent of cells, despite their robust detection by 4C (a 3C variant detecting all interactions with a specific bait sequence), suggesting that many chromosomal configurations are present within a population of cells ([Bantignies et al., 2011](#); [Noordermeer et al., 2011a](#); [Schoenfelder et al., 2010](#)). Despite this apparent diversity in global chromosome structure, several groups have attempted to model the average conformation (or conformations), which best globally fit the underlying interaction maps (for example [Duan et al., 2010](#); [Nagano et al., 2013](#); [Figure 4C](#)), whereas others have used more precise physical models to try and explain either the general features of Hi-C maps ([Barbieri et al., 2012](#); [Jost et al., 2014](#); [Lieberman-Aiden et al., 2009](#)) or obtain higher-resolution views of smaller genomic regions ([Giorgetti et al., 2014](#); [Le Dily et al., 2014](#)). More and higher-resolution interaction maps will allow the validity of these models to be tested, but already they have been able to provide testable hypotheses as to which genomic regions are the most crucial for structural integrity ([Giorgetti et al., 2014](#)).





**Figure 4. TAD-Dependent Enhancement of Chromatin Factor Targeting and Chromosome Conformation Heterogeneity**

(A) Top: A hypothetical TAD that contains three binding sites (in blue) for a chromatin factor is represented. Bottom: Intra-TAD contacts bring the chromatin binding sites in close proximity and form a 3D compartment where the chromatin factor is concentrated via formation of either homodimers or of self-interacting chromatin complexes. This architecture favors the maintenance of factor binding since, once the factor dissociates from a target site, the high relative concentration of other binding sites present in the same TAD favors rebinding.

(B) A genomic region with isolated binding sites for a chromatin factor (green) is shown. In the isolated context, the factor is rapidly lost in the nucleoplasm after dissociation from its target and therefore its replenishment from nucleoplasmic regions with lower relative concentration is less efficient. In this model, proposed by (Schuettengruber et al., 2014), 3D association of factor binding sites via intra-TAD contacts can favor the maintenance of robust chromatin targeting compared to non-TAD isolated factor binding sites.

(C) The tertiary structures of two mouse male T<sub>11.1</sub> cell X chromosomes, inferred from two separate single-cell Hi-C experiments, showing that single cells of a population can have diverse chromosome structures (Nagano et al., 2013). The chromosomal position of the fiber is shown as a color scale, going from red (centromeric end) to blue (telomeric end). The gray line represents regions with low constraints due to low mappability in the Hi-C experiment. Image provided by Csilla Varnai and Peter Fraser.

Comparisons of the chromatin interaction maps derived from multiple single-cell Hi-C experiments consistently revealed a high diversity in long-range contact repertoires but found that TADs were surprisingly persistent, suggesting that they are genuinely more stable sub-structures of the chromosome (Nagano et al., 2013; Figure 4C). What is currently unclear is how much of the structural heterogeneity is due to stable alternative genomic configurations and how much can be explained by chromosomal dynamics. Tagging mammalian DNA loci with multiple copies of binding sites for fluorescently labeled *lac* or *tet* repressors has revealed that chromatin is highly mobile but constrained within a restricted subnuclear volume (Lucas et al., 2014; Masui et al., 2011). This constrained diffusion is affected by developmental stage and attachment to nuclear landmarks

such as the periphery or nucleoli. On a larger scale, photobleaching studies of fluorescently labeled histones revealed that arrays of chromatin domains can undergo coordinated long-range movements (Cheutin and Cavalli, 2012). It is interesting to speculate that these domains could correlate with TADs (or groups of adjacent TADs), which have also been proposed to form the physical limit for the observed rapid sub-diffusion of chromatin (Lucas et al., 2014). Therefore, TADs may constitute the physical microenvironment in which neighboring functional elements interact, while occasional movements of strings of adjacent TADs may allow for large-scale rearrangement of chromosome structure and for the formation of new contacts among distant chromatin loci. A fascinating research area is to investigate whether these long-range movements might be specifically induced and regulated.

Moreover, very little is known about the conservation of chromosome structures across cell cycles; initial photobleaching experiments gave conflicting results for global chromosome positioning after mitosis (Gerlich et al., 2003; Walter et al., 2003). However, an elegant recent study suggests that at least some chromosome configurations can be remodeled during cell division. Lamina-associated chromatin was tagged during a short time period, and then its nuclear location(s) were traced through subsequent cell cycles (Kind et al., 2013). Only around one third of the lamina-associated chromatin called from population-average studies contacted the lamina at any given point in a single cell and, more strikingly, these regions were reshuffled during mitosis. Recent advances allow fluorescent DNA tagging without the insertion of large exogenous sequences (Chen et al., 2013; Miyazari et al., 2013; Saad et al., 2014). Their systematic application is likely to shed more light on the dynamics underpinning enhancer-promoter contacts, TAD stability and long-range interactions, and ultimately address whether they can be inherited across interphase and through subsequent cell cycles. Overall, whereas chromosomes are organized arrangements of seemingly stable secondary structures, they may adopt many different “tertiary structures” within a population, with as yet unclear dynamics of how these variants may interchange.

### Long-Range Interactions—Non-Opposites Attract

Focused 3C variants and FISH studies have uncovered a plethora of co-associations between genes separated by megabases, or occupying different chromosomes, usually occurring at frequencies that are low but much higher than expected by chance. Such long-range interactions are commonly between genes sharing regulation by a common factor, such as Polycomb-mediated repression (Bantignies et al., 2011; Denholtz et al., 2013), or activation by tissue-specific (Papantonis et al., 2012; Schoenfelder et al., 2010), or pluripotency-linked transcription factors (Apostolou et al., 2013; de Wit et al., 2013; Denholtz et al., 2013; Wei et al., 2013), occurring specifically in cell types where the regulation is mediated. Many groups have proposed the existence of functional spatial gene networks, whereby the clustering of genes at nuclear foci enriched in their regulatory factors facilitates their coordinate expression (Bantignies et al., 2011; Papantonis et al., 2012; Schoenfelder et al., 2010). Support for this model has come from detailed analysis of the acute co-association of three human TNF- $\alpha$

stimulated genes: an induced double-stranded DNA break in one gene completely abolishes its transcription but also severely impairs expression of the other target genes, concomitant with loss of co-association (Fanucchi et al., 2013). Most strikingly, this network is hierarchical, as break formation in the gene *SAMD4A* perturbs expression of both the genes *TNFAIP2* and *SLC6A5*, but *SAMD4A* is unaffected by breaks in either of the other genes. Similarly, a break in *TNFAIP2* perturbs *SLC6A5* expression but not vice versa. These examples of spatial co-regulated gene networks are very evocative; however in general, many combinations of genes sharing modes of regulation are not uncovered as interacting partners in 4C experiments. Furthermore, some gene co-associations linked to embryonic stem cell differentiation and formation of induced pluripotent cells precede the transcriptional changes by several days (Apostolou et al., 2013; Wei et al., 2013). It is also noteworthy that the observed spatial association of co-regulated genes in *S. cerevisiae* (Duan et al., 2010) was completely recapitulated when chromosomal structures were modeled from a few basic physical principles (Tjong et al., 2012). Thus, seemingly regulated spatial gene networks may actually be an indirect effect of chromosome folding mechanics, although the principles behind any potential direct regulation are even less clear than those determining enhancer-promoter communication or TAD organization at this stage.

Over multiple scales of chromosome organization, a recurring theme is the prevalence of homotypic or “like-with-like” interactions, whether this is the dimerization of proteins within chromatin loops (Deng et al., 2012), potential spatial networks of co-regulated genes (Schoenfelder et al., 2010) or a tendency for active and repressed chromatin to segregate (Lieberman-Aiden et al., 2009). Such configurations are the expected outcomes of self-organizing systems: a chance encounter between two loci bound by common regulatory factors increases the factors’ local concentrations, so that when a factor dissociates it is more likely to be re-trapped by the cluster of binding sites within its locale than to diffuse away to another location (Kang et al., 2011; Rajapakse et al., 2009). As association of the majority of DNA-bound factors with their cognate sites is transient (Phair and Misteli, 2000), self-organized spatial clustering of related genetic loci may be important for their efficient regulation. This model is consistent with the maintenance of active chromatin hubs at expressed genes (Palstra et al., 2003), the formation of Polycomb repressive domains (Lanzuolo et al., 2007), and perhaps their evolutionary robustness to motif mutations (Schuettengruber et al., 2014), and heterochromatic clustering (Taddei et al., 2009). As TAD organization mirrors underlying functional chromatin domains so well, we posit that TADs may be similarly self-organized structures that increase the local concentrations of diffusible regulatory factors around their sites of activity (Figures 4A and 4B). TADs may thus not only be an effective manner of preventing aberrant communication between genetic loci, but they may also allow for genes to be more efficiently bound by their effectors for stronger or more rapid transcriptional responses. Furthermore, the surprising finding that large-scale chromosome structures are actually more compact on perturbation of intra-TAD loops also suggests that TADs may be important for global chromosome structure maintenance (Tark-Dame et al., 2014).

## Perspectives

Mounting evidence shows that the genome is a dynamic yet highly organized hierarchical structure, built up from progressive stabilization of homotypic, potentially functional contacts between genes and regulatory elements. TADs present some conceptual analogy to secondary structures of proteins. These structures clearly have dynamics and cell-to-cell variability but also show a surprising developmental and evolutionary robustness, suggesting that they may be chromosome building blocks required for appropriate genome function. However, hypotheses about how TADs are organized and their functions are difficult to directly assess for two main reasons. First, up till now they have only been detected by population-average studies in fixed cells; TADs have yet to be visualized in single cells or followed in real time. Clearly, the way in which TADs may impinge on gene expression depends on whether they are genuinely stable structures or more a reflection of a ground state of inherent chromatin dynamics. Second, TADs appear robust to the initial perturbation studies that have been tried (for example, CTCF or cohesin abrogation), so it has been difficult to pinpoint any “causative” factor. Major advances in the future will tackle these two issues with live imaging of chromatin interactions in single cells (and following such interaction dynamics over the cell cycle), proteomic studies of which factors (if any) distinguish interacting loci from non-interacting ones and finer genetic dissection of the elements contributing to TAD borders or key architectural loops.

Returning to the protein folding analogy, genomes appear to be built up from the stabilization of progressively higher-order conformations, from TAD secondary structures to chromosomal tertiary structures, to the organized arrangement of chromosome territories into a final quaternary structure. With few exceptions, the structure of a protein cannot be predicted solely from its amino acid sequence. However, once the structure is resolved, the key residues contributing to the protein’s function can be readily identified and engineered. As our knowledge of TADs and specific chromatin loops increases, we posit that similar structure-informed reverse genetic engineering will allow us to manipulate the genome, with myriad applications. For example, de novo creation of autonomously regulated TADs would reduce any side effects of linked transgenes, and the engineering of switchable chromatin loops may allow for fine manipulation of gene expression. In summary, we are entering an exciting time in the field of nuclear organization. Mechanistic links are beginning to be assigned to what were previously only correlations between chromatin conformations and transcriptional regulation. Combined with the revolution in genome engineering tools such as CRISPR, we are in an unprecedented position to not only model, but also modulate, genome structure.

## ACKNOWLEDGMENTS

We thank Cyril Sarrauste de Menthier and Albert Weixlbaumer for help with the figures, and Csilla Varnai and Peter Fraser for contributing Figure 4C. Research at the G.C. lab is supported by grants from the European Research Council (ERC-2008-AdG No 232947), the CNRS, the European Network of Excellence EpiGeneSys, the Agence Nationale de la Recherche, and the Fondation ARC. Research at the T.S. lab is supported by INSERM, the ATIP-Avenir program, and the Laboratoire d’Excellence INRT.

## REFERENCES

- Apostolou, E., Ferrari, F., Walsh, R.M., Bar-Nur, O., Stadtfeld, M., Cheloufi, S., Stuart, H.T., Polo, J.M., Ohsumi, T.K., Borowsky, M.L., et al. (2013). Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* 12, 699–712.
- Ay, F., Bunnik, E.M., Varoquaux, N., Bol, S.M., Prudhomme, J., Vert, J.P., Noble, W.S., and Le Roch, K.G. (2014). Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.* 24, 974–988.
- Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144, 214–226.
- Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.M., Dostie, J., Pombo, A., and Nicodemi, M. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* 109, 16173–16178.
- Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. *Nat. Struct. Mol. Biol.* 20, 290–299.
- Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491.
- Cheutin, T., and Cavalli, G. (2012). Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genet.* 8, e1002465.
- Chiolo, I., Minoda, A., Colmenares, S.U., Polyzos, A., Costes, S.V., and Karpen, G.H. (2011). Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* 144, 732–744.
- Conaway, R.C., and Conaway, J.W. (2011). Function and regulation of the Mediator complex. *Curr. Opin. Genet. Dev.* 21, 225–230.
- Csank, A.K., and Henikoff, S. (1996). Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* 381, 529–531.
- de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502, 499–506.
- de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N., Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501, 227–231.
- DeMare, L.E., Leng, J., Cotney, J., Reilly, S.K., Yin, J., Sarro, R., and Noonan, J.P. (2013). The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* 23, 1224–1234.
- Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244.
- Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., and Blobel, G.A. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 158, 849–860.
- Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., Pellegrini, M., and Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* 13, 602–616.
- Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Drissen, R., Palstra, R.J., Gillemans, N., Splinter, E., Grosveld, F., Philipsen, S., and de Laat, W. (2004). The active spatial organization of the beta-globin locus requires the transcription factor EKLf. *Genes Dev.* 18, 2485–2490.
- Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. *Nature* 465, 363–367.
- Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shores, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43–49.
- Fanucchi, S., Shibayama, Y., Burd, S., Weinberg, M.S., and Mhlanga, M.M. (2013). Chromosomal contact permits transcription between coregulated genes. *Cell* 155, 606–620.
- Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol. Cell* 55, 694–707.
- Follmer, N.E., Wani, A.H., and Francis, N.J. (2012). A polycomb group protein is retained at specific sites on chromatin in mitosis. *PLoS Genet.* 8, e1003135.
- Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R., and Ellenberg, J. (2003). Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell* 112, 751–764.
- Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* 512, 96–100.
- Giorgetti, L., Galupa, R., Nora, E.P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 157, 950–963.
- Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol. Cell* 55, 678–693.
- Grzechnik, P., Tan-Wong, S.M., and Proudfoot, N.J. (2014). Terminate and make a loop: regulation of transcriptional directionality. *Trends Biochem. Sci.* 39, 319–327.
- Hadjur, S., Williams, L.M., Ryan, N.K., Cobb, B.S., Sexton, T., Fraser, P., Fisher, A.G., and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* 460, 410–413.
- He, B., Chen, C., Teng, L., and Tan, K. (2014). Global view of enhancer-promoter interactome in human cells. *Proc. Natl. Acad. Sci. USA* 111, E2191–E2199.
- Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. *Proc. Natl. Acad. Sci. USA* 109, 17507–17512.
- Heidari, N., Phanstiel, D.H., He, C., Grubert, F., Jahanbani, F., Kasowski, M., Zhang, M.Q., and Snyder, M.P. (2014). Genome-wide map of regulatory interactions in the human genome. *Genome Res.* 24, 1905–1917.
- Ho, J.W., Jung, Y.L., Liu, T., Alver, B.H., Lee, S., Ikegami, K., Sohn, K.A., Minoda, A., Tolstorukov, M.Y., Appert, A., et al. (2014). Comparative analysis of metazoan chromatin organization. *Nature* 512, 449–452.
- Hou, C., Dale, R., and Dean, A. (2010). Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. USA* 107, 3651–3656.
- Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* 48, 471–484.
- Ibn-Salem, J., Köhler, S., Love, M.I., Chung, H.R., Huang, N., Hurles, M.E., Haendel, M., Washington, N.L., Smedley, D., Mungall, C.J., et al. (2014). Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* 15, 423.
- Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294.

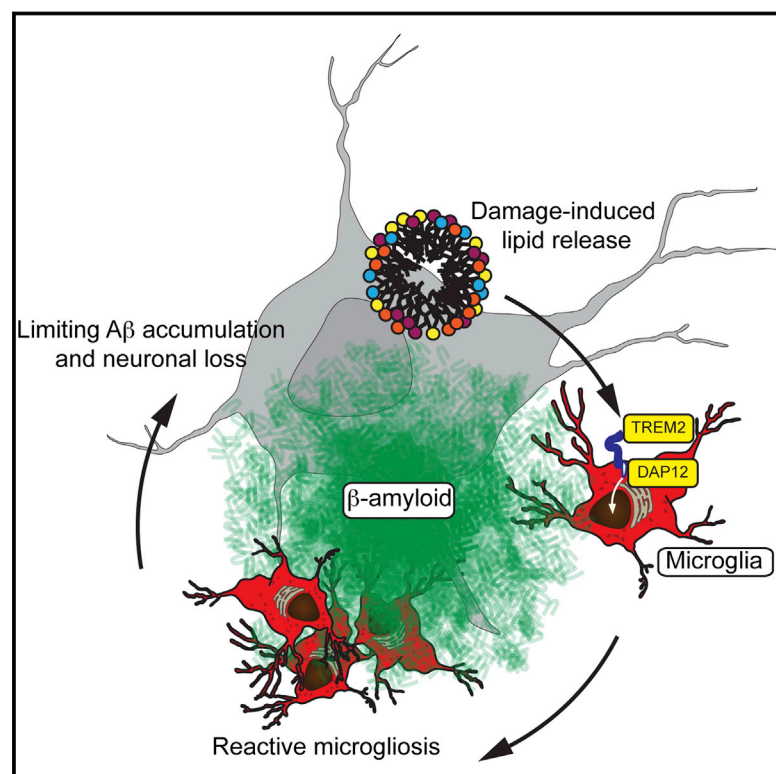
- Jost, D., Carrivain, P., Cavalli, G., and Vaillant, C. (2014). Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 42, 9553–9561.
- Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467, 430–435.
- Kang, J., Xu, B., Yao, Y., Lin, W., Hennessy, C., Fraser, P., and Feng, J. (2011). A dynamical model reveals gene co-localizations in nucleus. *PLoS Comput. Biol.* 7, e1002094.
- Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155, 1507–1520.
- Kind, J., Pagie, L., Ortabozkoyun, H., Boyle, S., de Vries, S.S., Janssen, H., Amendola, M., Nolen, L.D., Bickmore, W.A., and van Steensel, B. (2013). Single-cell dynamics of genome-nuclear lamina interactions. *Cell* 153, 178–192.
- Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. USA* 103, 10684–10689.
- Lanzuolo, C., Roue, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9, 1167–1174.
- Lavelle, C. (2014). Pack, unpack, bend, twist, pull, push: the physical side of gene expression. *Curr. Opin. Genet. Dev.* 25, 74–84.
- Le, T.B., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. *Science* 342, 731–734.
- Le Dily, F., Baù, D., Pohl, A., Vicent, G.P., Serra, F., Soronellas, D., Castellano, G., Wright, R.H., Ballare, C., Filion, G., et al. (2014). Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev.* 28, 2151–2162.
- Lemaître, C., Grabarz, A., Tsouroula, K., Andronov, L., Furst, A., Pankotai, T., Heyer, V., Rogier, M., Attwood, K.M., Kessler, P., et al. (2014). Nuclear position dictates DNA repair pathway choice. *Genes Dev.* 28, 2450–2463.
- Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148, 84–98.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Liu, M., and Grigoriev, A. (2004). Protein domains correlate strongly with exons in multiple eukaryotic genomes—evidence of exon shuffling? *Trends Genet.* 20, 399–403.
- Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* 158, 339–352.
- Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell* 145, 447–458.
- Miyanari, Y., Ziegler-Birling, C., and Torres-Padilla, M.E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* 20, 1321–1324.
- Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., and Grewal, S.I. (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432–435.
- Mousavi, K., Zare, H., Dell'orso, S., Grøntved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* 51, 606–617.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64.
- Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* 43, 525–558.
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. *Science* 342, 948–953.
- Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011a). Variegated gene expression caused by cell-specific long-range DNA interactions. *Nat. Cell Biol.* 13, 944–951.
- Noordermeer, D., Leleu, M., Splinter, E., Rougemont, J., De Laat, W., and Duboule, D. (2011b). The dynamic architecture of Hox gene clusters. *Science* 334, 222–225.
- Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385.
- Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35, 190–194.
- Papantonis, A., Kohro, T., Baboo, S., Larkin, J.D., Deng, B., Short, P., Tsutsumi, S., Taylor, S., Kanki, Y., Kobayashi, M., et al. (2012). TNF $\alpha$  signals through specialized factories where responsive coding and miRNA genes are transcribed. *EMBO J.* 31, 4404–4414.
- Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., et al. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* 38, 603–613.
- Phair, R.D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature* 404, 604–609.
- Phillips-Cremins, J.E., Sauria, M.E., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature* 515, 402–405.
- Rach, E.A., Winter, D.R., Benjamin, A.M., Corcoran, D.L., Ni, T., Zhu, J., and Ohler, U. (2011). Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.* 7, e1001274.
- Rajapakse, I., Perlman, M.D., Scalzo, D., Kooperberg, C., Groudine, M., and Kosak, S.T. (2009). The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. *Proc. Natl. Acad. Sci. USA* 106, 6679–6684.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Rubio, E.D., Reiss, D.J., Welch, P.L., Distèche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* 105, 8309–8314.
- Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-Gac, N., Lane, D., and Bystrycky, K. (2014). DNA dynamics during early double-strand break processing revealed by non-invasive imaging of living cells. *PLoS Genet.* 10, e1004187.
- Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature* 489, 109–113.



- Schauer, T., Schwalie, P.C., Handley, A., Margulies, C.E., Flicek, P., and Ladurner, A.G. (2013). CAST-ChIP maps cell-type-specific chromatin states in the *Drosophila* central nervous system. *Cell Rep.* 5, 271–282.
- Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42, 53–61.
- Schuettengruber, B., Oded Elkayam, N., Sexton, T., Entrevan, M., Stern, S., Thomas, A., Yaffe, E., Parrinello, H., Tanay, A., and Cavalli, G. (2014). Cooperativity, specificity, and evolutionary stability of Polycomb targeting in *Drosophila*. *Cell Rep.* 9, 219–233.
- Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-based chromatin interactions enable regulated gene expression within pre-existing architectural compartments. *Genome Res.* 23, 2066–2077.
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120.
- Sofueva, S., Yaffe, E., Chan, W.C., Georgopoulou, D., Vietri Rudan, M., Mira-Bontenbal, H., Pollard, S.M., Schroth, G.P., Tanay, A., and Hadjur, S. (2013). Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 32, 3119–3129.
- Soshnikova, N., and Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. *Science* 324, 1320–1323.
- Splinter, E., Heath, H., Kooren, J., Palstra, R.J., Klous, P., Grosveld, F., Galijart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354.
- Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J., Zhu, Y., Kaaij, L.J., van Ijcken, W., Gribnau, J., Heard, E., and de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev.* 25, 1371–1383.
- Symmons, O., Uslu, V.V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., Ettwiller, L., and Spitz, F. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400.
- Taddei, A., Van Houwe, G., Nagai, S., Erb, I., van Nimwegen, E., and Gasser, S.M. (2009). The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion. *Genome Res.* 19, 611–625.
- Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* 38, 8164–8177.
- Tark-Dame, M., Jerabek, H., Manders, E.M., Heermann, D.W., and van Driel, R. (2014). Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput. Biol.* 10, e1003877.
- Therizols, P., Illingworth, R.S., Courilleau, C., Boyle, S., Wood, A.J., and Bickmore, W.A. (2014). Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* 346, 1238–1242.
- Tjong, H., Gong, K., Chen, L., and Alber, F. (2012). Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* 22, 1295–1305.
- Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). *Drosophila* CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. *Genome Res.* 22, 2176–2187.
- Véron, A.S., Lemaitre, C., Gautier, C., Lacroix, V., and Sagot, M.F. (2011). Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics* 12, 303.
- Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., and Cremer, T. (2003). Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. *J. Cell Biol.* 160, 685–697.
- Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell Stem Cell* 13, 36–47.
- Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2014). Enhancer—core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559.
- Zaret, K.S. (2014). Genome reactivation after the silence in mitosis: recapitulating mechanisms of development? *Dev. Cell* 29, 132–134.
- Zhang, H., Jiao, W., Sun, L., Fan, J., Chen, M., Wang, H., Xu, X., Shen, A., Li, T., Niu, B., et al. (2013). Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. *Cell Stem Cell* 13, 30–35.
- Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654.
- Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van Ijcken, W.F., et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* 111, 996–1001.

# TREM2 Lipid Sensing Sustains the Microglial Response in an Alzheimer's Disease Model

## Graphical Abstract



## Authors

Yaming Wang, Marina Cella, ..., John R. Cirrito, Marco Colonna

## Correspondence

mcolonna@pathology.wustl.edu

## In Brief

TREM2 acts in microglia as a sensor for a wide array of lipids that are associated with  $\beta$ -amyloid accumulation and neuronal loss. The TREM2 mutation that has recently been identified as a risk factor for Alzheimer's disease attenuates microglial detection of damage-associated lipids, providing a mechanistic basis for the genetic association.

## Highlights

- TREM2 deficiency augments A $\beta$  accumulation and neuronal loss in a mouse model of AD
- TREM2 sustains the microglial response to A $\beta$  plaques by promoting microglial survival
- TREM2 senses anionic lipids that have been found to interact with fibrillar A $\beta$
- TREM2 R47H mutation linked to Alzheimer's disease impairs lipid recognition

## Accession Numbers

GSE65067



Wang et al., 2015, Cell 160, 1061–1071  
March 12, 2015 ©2015 Elsevier Inc.  
<http://dx.doi.org/10.1016/j.cell.2015.01.049>

# TREM2 Lipid Sensing Sustains the Microglial Response in an Alzheimer's Disease Model

Yaming Wang,<sup>1,5</sup> Marina Cella,<sup>1</sup> Kaitlin Mallinson,<sup>2,3,4</sup> Jason D. Ulrich,<sup>2,3,4</sup> Katherine L. Young,<sup>2,3,4</sup> Michelle L. Robinette,<sup>1</sup> Susan Gilfillan,<sup>1</sup> Gokul M. Krishnan,<sup>1</sup> Shwetha Sudhakar,<sup>2,3,4</sup> Bernd H. Zinselmeyer,<sup>1</sup> David M. Holtzman,<sup>2,3,4</sup> John R. Cirrito,<sup>2,3,4</sup> and Marco Colonna<sup>1,\*</sup>

<sup>1</sup>Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

<sup>2</sup>Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA

<sup>3</sup>Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis, MO 63110, USA

<sup>4</sup>Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO 63110, USA

<sup>5</sup>Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA

\*Correspondence: [mcolonna@pathology.wustl.edu](mailto:mcolonna@pathology.wustl.edu)

<http://dx.doi.org/10.1016/j.cell.2015.01.049>

## SUMMARY

Triggering receptor expressed on myeloid cells 2 (TREM2) is a microglial surface receptor that triggers intracellular protein tyrosine phosphorylation. Recent genome-wide association studies have shown that a rare R47H mutation of TREM2 correlates with a substantial increase in the risk of developing Alzheimer's disease (AD). To address the basis for this genetic association, we studied TREM2 deficiency in the 5XFAD mouse model of AD. We found that TREM2 deficiency and haploinsufficiency augment  $\beta$ -amyloid (A $\beta$ ) accumulation due to a dysfunctional response of microglia, which fail to cluster around A $\beta$  plaques and become apoptotic. We further demonstrate that TREM2 senses a broad array of anionic and zwitterionic lipids known to associate with fibrillar A $\beta$  in lipid membranes and to be exposed on the surface of damaged neurons. Remarkably, the R47H mutation impairs TREM2 detection of lipid ligands. Thus, TREM2 detects damage-associated lipid patterns associated with neurodegeneration, sustaining the microglial response to A $\beta$  accumulation.

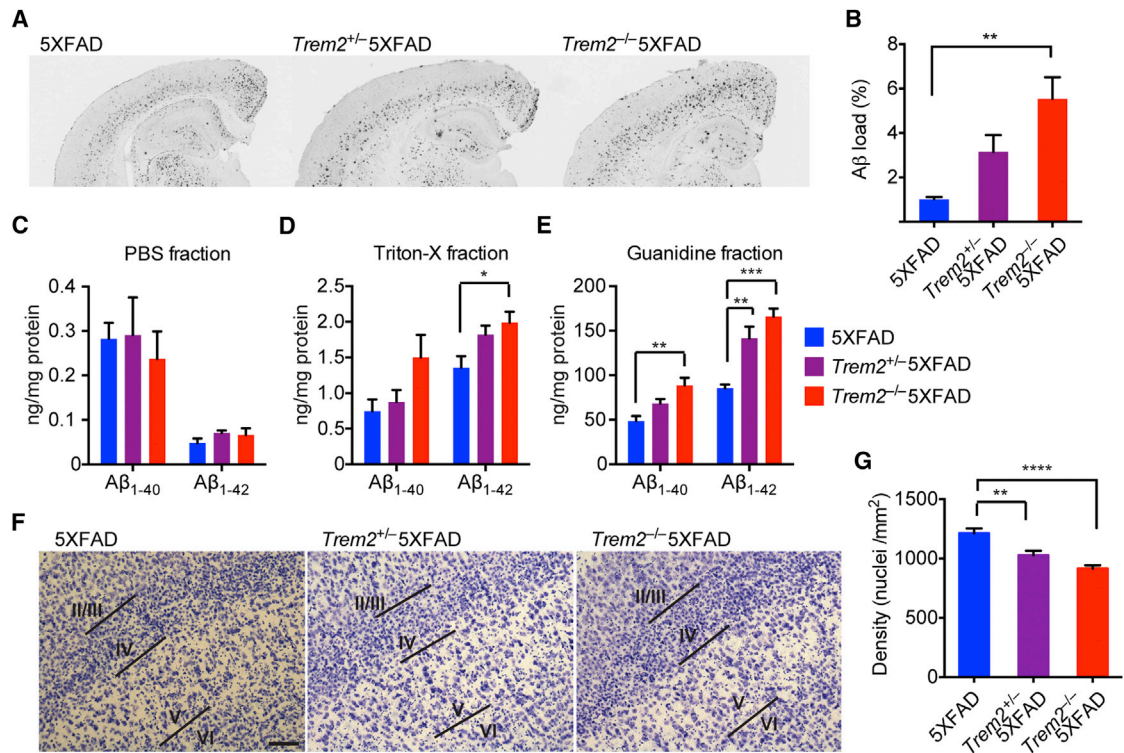
## INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder with histopathological hallmarks of  $\beta$ -amyloid (A $\beta$ ) plaques and neurofibrillary tangles in the brain (Huang and Mucke, 2012; Tanzi, 2013). Although disease etiology is incompletely understood, families with inherited early-onset AD have mutations in three proteins directly involved in the A $\beta$  processing pathway, suggesting a key role for A $\beta$  in disease pathogenesis. Early studies have shown that brain microglia accumulate around A $\beta$  plaques and occasionally contain A $\beta$  in both AD patients (D'Andrea et al., 2004; McGeer et al., 1987; Perlmuter et al., 1990) and transgenic mouse models of AD (Dickson, 1999; Frautschy et al., 1998; Stalder et al., 1999). Microglia contribute to A $\beta$  clearance, at least in the early phases of neurodegeneration (El Khoury et al.,

2007); however, the ability of microglia to clear A $\beta$  may wane with age (Streit et al., 2004; Streit and Xue, 2009). At late stages of AD, microglia may paradoxically contribute to the disease by releasing pro-inflammatory cytokines in response to A $\beta$  deposition (El Khoury et al., 2007; Hickman et al., 2008).

Recent genome-wide association studies (GWASs) have shown that a rare Arginine-47-Histidine (R47H) mutation of the triggering receptor expressed on myeloid cells 2 (TREM2) is associated with a substantial increase in the risk of developing AD (Guerreiro et al., 2013b; Jonsson et al., 2013). TREM2 is a cell-surface receptor of the Ig-superfamily that is expressed by microglia and osteoclasts in vivo (Kilalainen et al., 2005; Paloneva et al., 2002; Schmid et al., 2002; Thrash et al., 2009) as well as monocyte-derived DCs, bone marrow-derived macrophages, and macrophage cell lines in vitro (Bouchon et al., 2001; Daws et al., 2001). Although TREM2 was detected in other cells of the CNS (Guerreiro et al., 2013b; Sessa et al., 2004), these observations have not been confirmed (Jiang et al., 2014). TREM2 binds anionic carbohydrates, anionic bacterial products, and various phospholipids (Cannon et al., 2012; Daws et al., 2003). It transmits intracellular signals through the associated transmembrane adaptor DAP12, which recruits the protein tyrosine kinase Syk, leading to phosphorylation of many downstream mediators, such as PLC- $\gamma$ , PI-3K, and Vav2/3 (Ford and McVicar, 2009; Peng et al., 2010). Individuals homozygous for rare mutations that impair expression of either TREM2 or DAP12 develop lethal forms of progressive dementias such as Nasu-Hakola disease (NHD) and frontotemporal dementia (FTD) (Guerreiro et al., 2013a, 2013c; Kleinberger et al., 2014; Paloneva et al., 2002).

The association between the R47H mutation of TREM2 and the increased risk for late-onset AD suggests that microglia may require TREM2 to respond to A $\beta$  deposition and to limit neuronal degeneration. Consistent with this hypothesis, we recently showed that APPPS1-21 transgenic mice, an AD model with rapid deposition of A $\beta$ , have a marked decrease in the number and size of A $\beta$ -associated microglia when they lack one copy of the *Trem2* gene, although this defect did not increase A $\beta$  accumulation (Ulrich et al., 2014). The mechanisms underlying this altered microglial response and its impact on A $\beta$  deposition have not been delineated. To address these questions, we studied TREM2 deficiency in the 5XFAD mouse model of AD, in which



**Figure 1. TREM2-Deficient 5XFAD Mice Have Increased Hippocampal Aβ Burden and Accelerated Loss of Layer V Cortical Neurons**

Aβ burden in 8.5-month-old *Trem2*<sup>-/-</sup>5XFAD, *Trem2*<sup>+/-</sup>5XFAD, and 5XFAD mice.

(A) Matching coronal hippocampus and cortex sections were stained with an Aβ-specific antibody mHJ3.4.

(B) Amounts of Aβ loads in hippocampi.

(C–E) Soluble and insoluble Aβ<sub>1-40</sub> and Aβ<sub>1-42</sub> levels in hippocampi as detected by ELISA.

(F and G) Densities of layer V neurons in 8.5-month-old *Trem2*<sup>-/-</sup>5XFAD, *Trem2*<sup>+/-</sup>5XFAD, and 5XFAD mice. (F) Matching coronal sections stained with cresyl violet. (G) Summary of densities of layer V neurons.

Original magnification: 10×; scale bar, 100 μm. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, \*\*\*\*p < 0.0001, one-way ANOVA. Data represent analyses total of eight to ten 5XFAD mice, eight to 12 *Trem2*<sup>+/-</sup>5XFAD mice, and eight to 16 *Trem2*<sup>-/-</sup>5XFAD mice (B, C–E, and G). Bars represent mean ± SEM. See also Figure S1.

Aβ deposition develops less rapidly than in APPS1-21 mice (Oakley et al., 2006). We find that both TREM2 deficiency and haploinsufficiency augment Aβ accumulation due to a dysfunctional response of microglia, which become apoptotic rather than undergoing activation and proliferation. We further show that TREM2 sustains microglial survival by synergizing with colony stimulating factor-1 receptor (CSF-1R) signaling. Finally, we demonstrate that TREM2 binds to a broad array of anionic lipids, which were found in association with fibrillar Aβ and are also exposed during neuronal and glial cell death. Remarkably, the R47H mutation impairs TREM2 binding to anionic lipids. We conclude that TREM2 is a receptor that detects damage-associated lipids, thereby enabling microglia to sense Aβ accumulation and cell damage, as well as supporting microglial survival and Aβ reactive microgliosis.

## RESULTS

### TREM2 Modulates Aβ Accumulation

We examined the deposition of Aβ aggregates in *Trem2*<sup>-/-</sup> mice bred to 5XFAD transgenic mice (APPSwF10n, PSEN1\*M146L\*L286V), an accelerated mouse model of AD (Oakley et al.,

2006). Staining of matched coronal brain sections from *Trem2*<sup>-/-</sup>5XFAD mice and control 5XFAD mice at 8.5 months of age with a monoclonal antibody (mAb) against Aβ revealed significantly increased Aβ accumulation in the hippocampus but not cortical regions of *Trem2*<sup>-/-</sup>5XFAD mice (Figures 1A, 1B, and S1A). *Trem2*<sup>+/-</sup>5XFAD mice had an intermediate phenotype, although it was not statistically significant (p = 0.104). We also determined levels of Aβ<sub>40</sub> and Aβ<sub>42</sub> in the hippocampus and cortex of these mice by ELISA. While levels of soluble Aβ<sub>40</sub> and Aβ<sub>42</sub> were similar (Figures 1C and S1B), we detected a significant increase in insoluble, guanidine-extracted Aβ<sub>40</sub> and Aβ<sub>42</sub> in the hippocampal regions of *Trem2*<sup>-/-</sup>5XFAD mice compared to 5XFAD mice (Figures 1D and 1E). Moreover, there was a significant effect of *Trem2* gene copy number on insoluble Aβ protein levels in the hippocampi, whereas levels of insoluble Aβ<sub>40</sub> and Aβ<sub>42</sub> in the cortex were equivalent across all three genotypes (Figures S1C and S1D). We also found that the loss of layer V neurons, a feature of 5XFAD mice (Eimer and Vassar, 2013; Oakley et al., 2006), was more prominent in *Trem2*<sup>-/-</sup>5XFAD mice (Figures 1F and 1G). *Trem2*<sup>+/-</sup>5XFAD mice presented an intermediate phenotype. Collectively, these data suggest that TREM2 modulates Aβ accumulation, limiting neuronal



loss. The lack of a significant difference in A $\beta$  accumulation in the cortices of *Trem2*<sup>-/-</sup>5XFAD mice and 5XFAD mice may be the result of the fast kinetics of A $\beta$  deposition in 5XFAD mice, such that the potential cortical differences are no longer detectable at 8.5 month of age.

### TREM2 Is Required for Reactive Microgliosis

How does lack of TREM2 impact A $\beta$  accumulation? Although TREM2 expression has been reported in CNS cells other than microglia (Guerreiro et al., 2013b; Sessa et al., 2004), this finding is controversial (Jiang et al., 2014). Indeed, a recently published RNA sequencing (RNA-seq) data set demonstrated that *Trem2* is specifically expressed in microglia, but not other cells in the CNS under steady-state conditions (Butovsky et al., 2014). We also found that *Trem2* expression is further upregulated in microglia isolated from 5XFAD mice during A $\beta$  deposition (Figures S2A and S2B). Thus, we focused our studies on microglia. One of the many effects of A $\beta$  deposition is the induction of reactive microgliosis, which involves the expansion of microglia and conversion to an activated state (Ransohoff and Cardona, 2010). Microgliosis predominantly involves the proliferation of brain-resident microglia, with some contribution from blood-borne monocytes and microglia migrating from adjacent non-damaged brain areas (El Khoury et al., 2007; Grathwohl et al., 2009; Malm et al., 2005; Mildner et al., 2011; Simard et al., 2006; Stalder et al., 2005). To evaluate the impact of TREM2 deficiency on A $\beta$ -induced microglial responses in 5XFAD mice, we examined transcriptional profiles of microglia purified from 5XFAD and *Trem2*<sup>-/-</sup>5XFAD mice as well as transgene negative wild-type (WT) and *Trem2*<sup>-/-</sup> littermates (Figure S2A). To evaluate changes in global transcripts, we first performed principle component analysis (PCA) of the top 15% most variable transcripts. We noticed that WT and *Trem2*<sup>-/-</sup> replicates clustered closely, suggesting a limited impact of TREM2 deficiency in the steady state, which was confirmed by a volcano plot comparing the two groups (Figures 2A and 2B). In contrast, 5XFAD microglial replicates were dramatically different from WT replicates (Figure 2A), and a volcano plot revealed that 5XFAD microglia expressed many more transcripts including those associated with microglial activation (MHC-II, CD11c), production of inflammatory cytokines (interleukin-1 $\beta$  [IL-1 $\beta$ ], tumor necrosis factor- $\alpha$  [TNF- $\alpha$ ], IL-12, and SPP1), and neurotrophic factors (insulin growth factor 1 [IGF-1] and VEGFA) (Figure 2C). *Trem2*<sup>-/-</sup>5XFAD microglia had an intermediate behavior in the principle component analysis compared to 5XFAD and WT microglia. To further interrogate how TREM2 deficiency affected the microglial response to A $\beta$  deposition, we selected the transcripts upregulated 2-fold between 5XFAD and WT microglia (Figure 2C) and compared the expression of these transcripts among the entire data set. We found that *Trem2*<sup>-/-</sup>5XFAD microglia failed to upregulate these transcripts and behaved more similarly to WT microglia, as shown by hierarchical clustering and expression-by-expression plots (Figures 2D and 2E). Flow cytometric analysis of isolated microglia confirmed phenotypic changes in 5XFAD microglia consistent with increased activation, including a marked increase in cell size and strong upregulation of MHC-II, CD11c, and CD11b (Figures S2C–S2G). We also confirmed

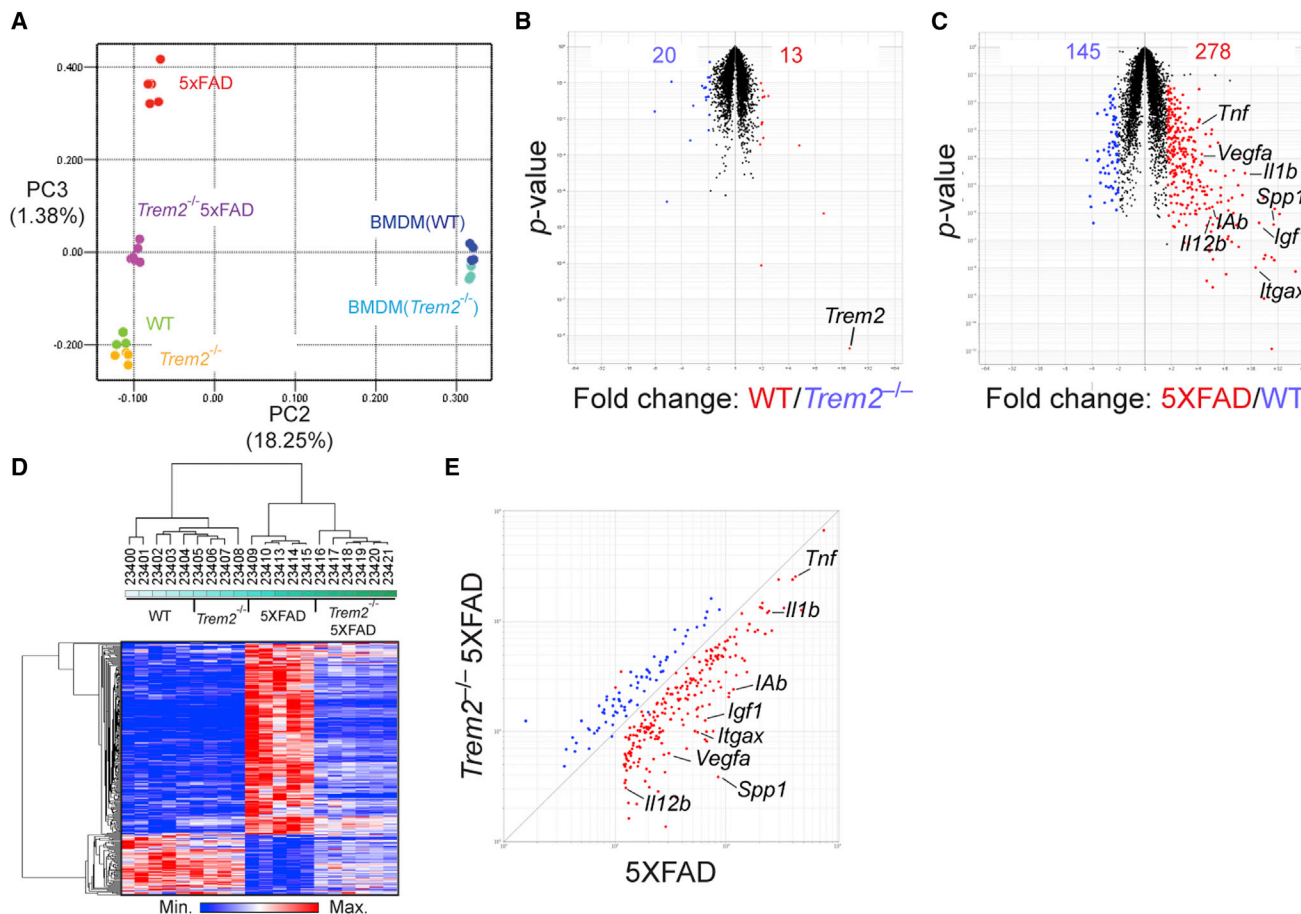
increased expression of inflammatory cytokine transcripts by qPCR in whole-brain lysates of 5XFAD mice (Figures S2I–S2L). However, in *Trem2*<sup>-/-</sup>5XFAD mice, these changes were markedly attenuated (Figures 2D, 2E, and S2C–S2L). In fact, *Trem2*<sup>-/-</sup>5XFAD microglia were phenotypically more similar to WT microglia in steady state than 5XFAD microglia. Overall, these results implied that TREM2 is required for reactive microgliosis.

### Microglia Fail to Colocalize with A $\beta$ Plaques in *Trem2*<sup>-/-</sup> Mice

Initial staining of microglia in coronal brain sections with Iba-1 revealed very similar distribution of microglia in *Trem2*<sup>-/-</sup>, *Trem2*<sup>+/-</sup> and WT adult mice (Figures S3A–S3C). However, co-staining of coronal brain sections from *Trem2*<sup>-/-</sup>5XFAD and 5XFAD mice with Iba-1 and X-34, to visualize microglia and A $\beta$  plaques, respectively, showed remarkable differences. We found that *Trem2*<sup>-/-</sup>5XFAD mice had reduced Iba-1 reactivity both in the hippocampi and cortices compared to 5XFAD mice (Figures 3A–3D). This was particularly evident in the areas surrounding A $\beta$  plaques (Figures 3E and 3F; Movies S1, S2, and S3), suggesting a preferential reduction of microgliosis near amyloid deposits. *Trem2*<sup>+/-</sup>5XFAD mice also had a partial reduction of amyloid-associated Iba-1 reactivity.

Examination of a second model of AD, APPPS1-21 mice that have been bred to *Cx3cr1*<sup>GFP/+</sup> mice in order to visualize endogenous microglia, confirmed that complete TREM2 deficiency results in a marked reduction of GFP<sup>+</sup> microglial clusters around A $\beta$  plaques (Figures S3D–S3F). This corroborates our previous observation that TREM2 haploinsufficiency correlates with fewer amyloid-associated microglia in APPPS1-21x*Cx3cr1*<sup>GFP/+</sup> mice (Ulrich et al., 2014). Moreover, since CX3CR1 marks brain-resident microglia (Ransohoff and Cardona, 2010), these results also suggest that TREM2 deficiency primarily affects the response of brain-resident microglia to A $\beta$ .

To further quantify the number of microglia around A $\beta$  plaques, we recorded the coordinates (x, y, and z) of all visible microglial cell bodies and the location of A $\beta$  plaques in each z stack confocal image and calculated the number of microglia within 30  $\mu$ m radius of the plaques (defined as plaque-associated microglia) and non-plaque-associated microglia. While no statistically significant difference was observed among non-plaque-associated microglia (Figure S4A), we noted a high degree of microglial clustering around amyloid plaques in 5XFAD mice (average 4.28 microglia per plaque), which gradually decreased in *Trem2*<sup>+/-</sup>5XFAD mice (average 3.42 microglia per plaque) and *Trem2*<sup>-/-</sup>5XFAD mice (average 2.36 microglia per plaque) (Figures 4A and 4B). To confirm the “negligence” of microglial responses to A $\beta$  in the absence of TREM2, we compared the actual frequency of microglia per plaque to that obtained by Monte Carlo simulations where the same numbers of microglia and plaques observed in z stack images were positioned by chance in each genotype (Figure S4B). The probability that observed microglial frequencies per plaque fell outside of simulated random frequencies was inversely proportional to *Trem2* gene copy number (Figure 4C). Moreover, while 27.9% of microglial distribution in 5XFAD mice with respect to A $\beta$  plaques was



**Figure 2. TREM2 Deficiency Impairs A $\beta$ -Induced Transcriptional Program in Microglia**

Transcriptional analysis of microglia isolated from hippocampi and cortices of 8.5-month-old *Trem2*<sup>-/-</sup>5xFAD, 5xFAD, *Trem2*<sup>-/-</sup>, and WT mice.

(A) Top 15% most variable transcripts were subjected to principle component analysis (PCA). Plot shows two-dimensional (PC2 versus PC3) comparison of transcriptional changes in all classes analyzed. WT and *Trem2*<sup>-/-</sup> bone marrow-derived macrophages were used as references.

(B) Volcano plot comparing microglial transcripts in *Trem2*<sup>-/-</sup> and WT mice. *Trem2* transcript is indicated.

(C) Volcano plot comparing microglial transcripts in 5xFAD and WT mice. Numbers in plots (B) and (C) indicate probes that are significantly upregulated or downregulated ( $\pm 2$ -fold,  $p < 0.05$ , Student's *t* test). Representative transcripts are indicated.

(D and E) Visualization of A $\beta$ -induced changes in microglial transcripts from (C). (D) A heatmap displays hierarchical clustering of all samples analyzed. (E) A scatterplot compares these transcriptional changes in *Trem2*<sup>-/-</sup>5xFAD and 5xFAD microglia. Representative transcripts are shown.

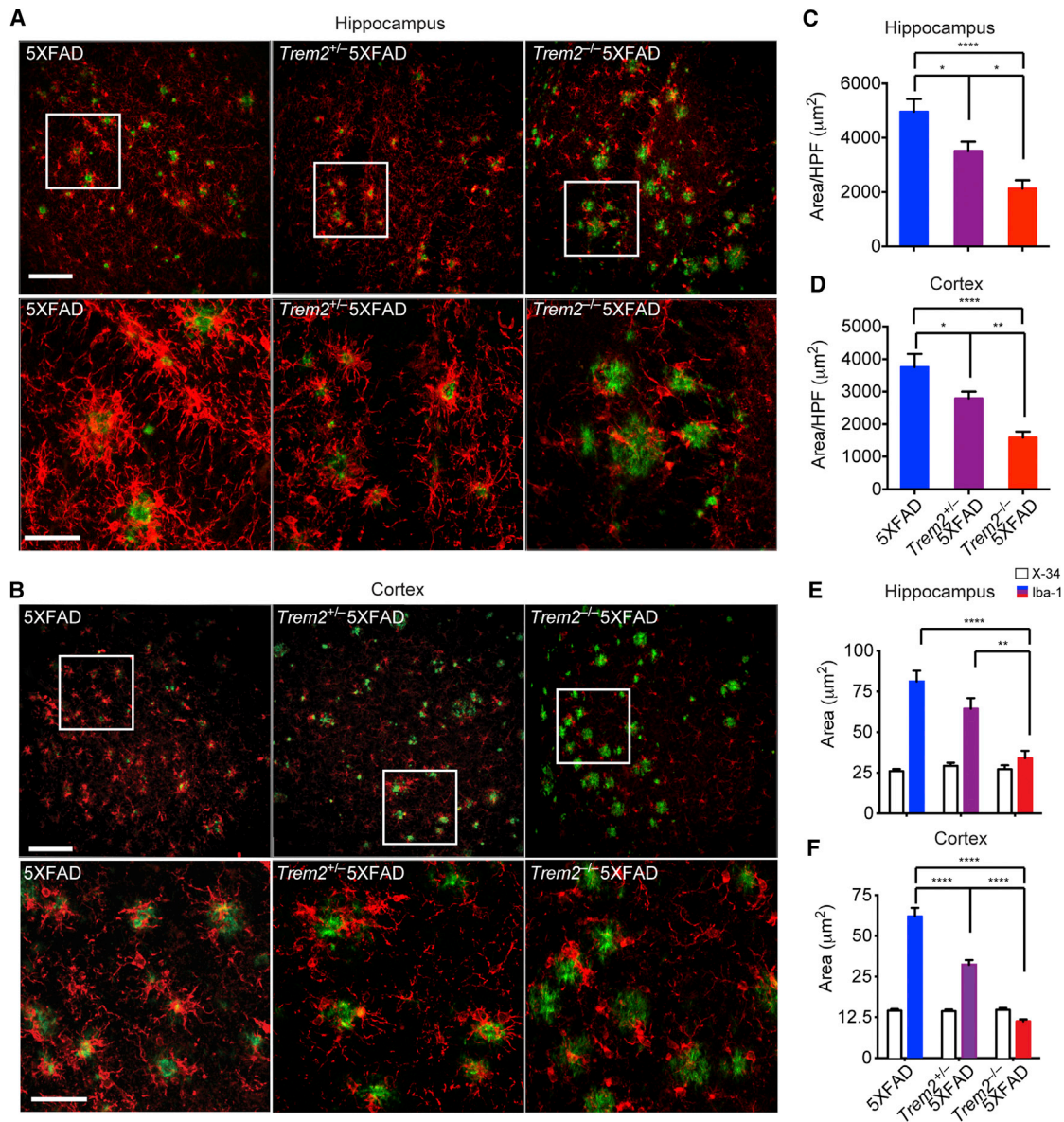
See also Figure S2.

not explained statistically by chance, the frequency of non-random microglial distribution was reduced to 9.5% in *Trem2*<sup>-/-</sup> 5xFAD mice (Figure 4D).

Another feature of reactive microgliosis is morphological transformation. In 5xFAD mice, plaque-associated microglia showed morphological changes associated with microglial activation, including a partial retraction and a slight hypertrophy of the microglial cell processes as well as an increase in size (Figures 4E–4G). These changes in microglial morphology were significantly attenuated in *Trem2*<sup>+/-</sup>5xFAD and *Trem2*<sup>-/-</sup> 5xFAD mice (Figures 4E–4G) and were paralleled by an increased distance between microglia and the center of their associated plaques (Figure 4H). Collectively, these data indicate that TREM2 is essential for the microglial response to A $\beta$  plaques.

### TREM2 Deficiency Affects Microglial Survival in 5xFAD Mice

Why is TREM2 required for A $\beta$  reactive microgliosis? We first hypothesized that TREM2 may be necessary for A $\beta$  uptake and microglial activation. We initially investigated the impact of TREM2 deficiency on microglial activation in vitro. For this analysis, we used primary microglia isolated from adult mice and expanded in the presence of optimal amounts of CSF-1 and TGF- $\beta$  (Figure S5A), as they closely resemble microglia in vivo (Butovsky et al., 2014). TREM2 deficiency did not affect microglial expansion, migration, or TNF- $\alpha$  secretion in response to A $\beta$  (Figures S5B–S5D). In contrast, *Trem2*<sup>-/-</sup> microglia produced significantly more TNF- $\alpha$  than WT microglia in response to lipopolysaccharide (LPS), consistent with previous demonstrations that TREM2 attenuates cytokine responses to certain TLR ligands



**Figure 3. TREM2 Deficiency Leads to Reduced Microgliosis in 5XFAD Mice**

Microgliosis in 8.5-month-old *Trem2*<sup>+/-</sup>5XFAD, *Trem2*<sup>-/-</sup>5XFAD, and 5XFAD mice.

(A and B) Matching coronal sections were stained with Iba-1 (red) for microglia and X-34 (green) for amyloid plaques. Representative z stack images with maximum projection are shown.

(C and D) Quantification of total Iba-1 reactivity per high-power field (HPF) in hippocampi and cortices.

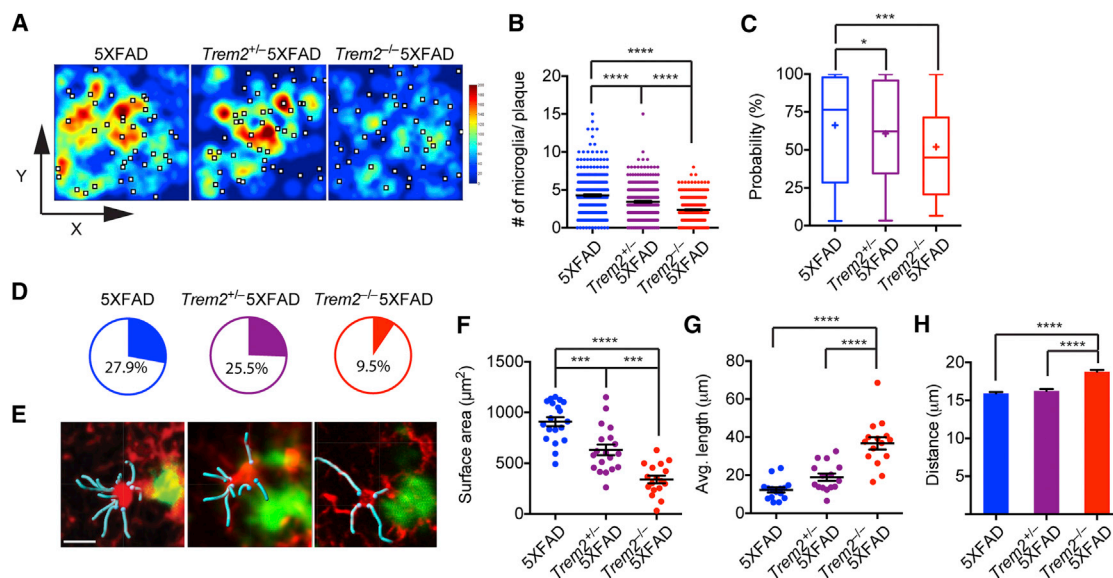
(E and F) Quantification of microgliosis associated with plaques of similar sizes in hippocampi and cortices.

Original magnification 20× (A, B, upper panels), 40× (A, B, lower panels); scale bar, 10 μm (A, B, upper panels), 50 μm (A, B, lower panels). \*p < 0.05, \*\*p < 0.01, \*\*\*\*p < 0.0001, one-way ANOVA. Data represent analyses of a total of eight of ten 5XFAD, eight of 12 *Trem2*<sup>+/-</sup>5XFAD mice, and eight of 16 *Trem2*<sup>-/-</sup>5XFAD mice. Bars represent mean ± SEM. See also Figure S3 and Movies S1, S2, and S3.

(Hamerman et al., 2006; Turnbull et al., 2006). Moreover, TREM2 deficiency had very little impact on microglial uptake of Aβ aggregates (Figure S5E; Movie S4) or their subsequent proteolytic processing, as demonstrated by similar degradation of the intracellular concentration of Aβ after initial loading (Figure S5F). Thus, TREM2 deficiency does not engender a direct defect in phagocytosis of Aβ.

Previous studies have suggested that the CSF-1-CSF-1R pathway promotes reactive microgliosis (Chitu and Stanley, 2006) and Aβ clearance (Mitrassinovic et al., 2003); consistent with this, CSF1-deficient osteopetrotic (*op/op*) mice are characterized by increased deposition of Aβ, scarcity of microgliosis and neuronal loss (Kaku et al., 2003). We had previously demonstrated that TREM2 signaling via its associated adaptor DAP12





**Figure 4. TREM2 Deficiency Diminishes the Capacity of Microglia to Cluster around A $\beta$  Plaques**

Frequencies of plaque-associated microglia in 8.5-month-old *Trem2*<sup>-/-</sup>5XFAD, *Trem2*<sup>+/-</sup>5XFAD, and 5XFAD mice were determined.

(A) Heatmap shows frequencies of microglia in relation to A $\beta$  plaques shown as white squares.

(B) Summary of frequencies of plaque-associated microglia in all analyzed genotypes.

(C and D) Microglial clustering around plaques in 5XFAD, *Trem2*<sup>+/-</sup>5XFAD, and *Trem2*<sup>-/-</sup>5XFAD mice were compared to Monte Carlo simulations that assume total randomness between plaques and microglia. Probabilities that any given microglia-plaque cluster are non-random are shown in (C). Pie charts show frequencies of microglia-plaque clusters that cannot be statistically explained as random ( $p < 0.05$ ) (D).

(E) Morphology of plaque-associated microglia highlighting the shape of cell bodies (red) and primary processes (cyan).

(F–H) Plaque-associated microglia are analyzed for their surface area (cell body only), average length of primary processes, and distance from the center of adjacent A $\beta$  plaque.

Original magnification: 20 $\times$ ; scale bar, 15  $\mu$ m. \* $p < 0.05$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ , one-way ANOVA. Data represent analyses of a total seven mice per group (A–D) and a total of five mice per group (E–G). Bars represent mean  $\pm$  SEM. See also Figure S4.

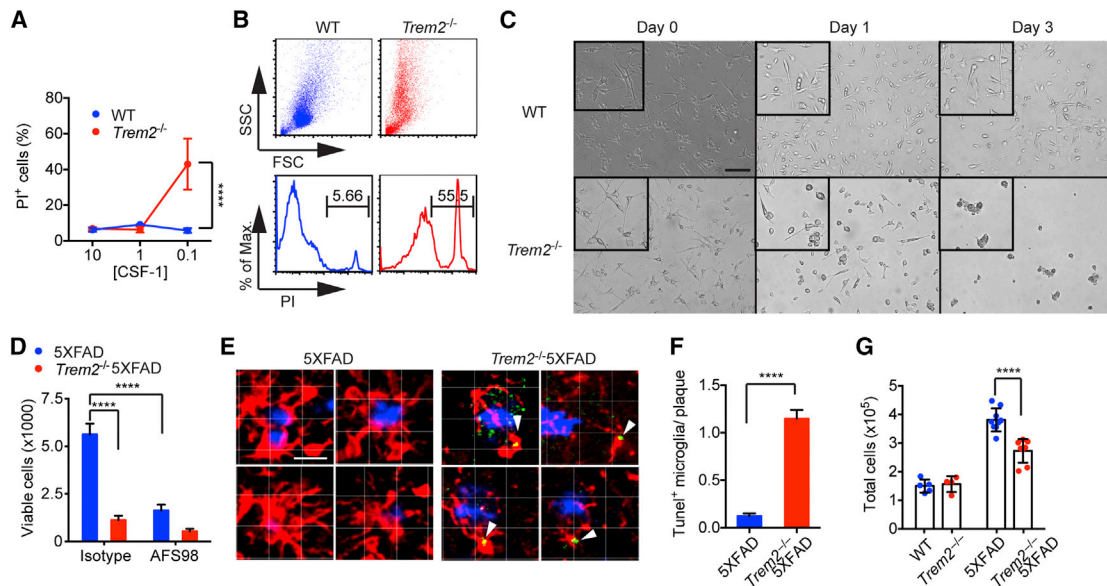
synergizes with CSF-1R signaling to promote survival of macrophages (Otero et al., 2009, 2012). Specifically, TREM2/DAP12 were required to induce activation of the Syk tyrosine kinase pathway downstream of CSF-1R (Otero et al., 2009; Zou et al., 2008). Thus, we hypothesized that TREM2 may synergize with CSF-1-CSF-1R signaling to sustain reactive microgliosis during A $\beta$  deposition. We initially tested this hypothesis in vitro by measuring the survival of adult primary microglial cultures from WT and *Trem2*<sup>-/-</sup> mice in the presence of graded concentrations of CSF-1 (10%, 1%, and 0.1% L-cell conditioned medium [LCM]). While TREM2 deficiency did not affect viability at high concentrations of CSF-1 (10% and 1%), *Trem2*<sup>-/-</sup> microglia were markedly less viable than WT microglia in 0.1% CSF-1 (Figures 5A–5C). We next purified microglia from *Trem2*<sup>-/-</sup>5XFAD and 5XFAD mice and cultured them in medium containing low levels of CSF-1 (0.1% LCM) for 5 days. *Trem2*<sup>-/-</sup>5XFAD microglia were significantly less viable than 5XFAD microglia (Figure 5D). Since CSF-1R captures CSF-1 and targets it for degradation (Stanley and Chitu, 2014), the reduced survival of *Trem2*<sup>-/-</sup> microglia at low CSF-1 concentrations may reflect a marked susceptibility of these cells to CSF-1 deprivation that occurs when microglia consume a limited supply of CSF-1. Indeed, CSF-1R blockade reduced viability of 5XFAD microglia, confirming that the pro-survival effect of TREM2 cannot replace that of CSF-1R, but only synergize with it (Figure 5D).

To evaluate the impact of TREM2 deficiency on microglia apoptosis in vivo, we analyzed coronal sections of *Trem2*<sup>-/-</sup>5XFAD and 5XFAD mice by TUNEL staining. Markedly more TUNEL<sup>+</sup> microglia were evident in *Trem2*<sup>-/-</sup>5XFAD mice than the very few observed in control 5XFAD mice (Figures 5E and 5F), corroborating a role for TREM2 in maintaining microglial survival during reactive microgliosis. Consistent with this, significantly fewer microglia were recovered from the cortices and hippocampi of *Trem2*<sup>-/-</sup>5XFAD mice than from 5XFAD mice (Figure 5G). We postulate that reactive microgliosis is associated with increased CSF-1 uptake by CSF-1R and degradation restricting CSF-1 range of action, such that microglia in close proximity must compete for CSF-1. Because of their inability to survive CSF-1 limitation, TREM2-deficient microglia are incapable of sustaining reactive microgliosis and undergo apoptosis rather than becoming activated and expanding.

#### **TREM2 Is a Sensor for Anionic and Zwitterionic Lipids that Accumulate in the CNS during A $\beta$ Deposition**

We next sought to identify the ligand(s) that trigger TREM2 signaling during A $\beta$  deposition. Since TREM2 binds anionic carbohydrates, anionic bacterial products, and phospholipids (Cannon et al., 2012; Daws et al., 2003), we focused on lipids that have been shown to accumulate during A $\beta$  deposition and





**Figure 5. TREM2 Promotes Microglial Survival Ex Vivo and In Vivo**

(A–C) Adult primary microglia were cultured with various concentration of CSF-1-containing L-cell medium (LCM). Viability of microglia by PI staining (A and B) and morphology (C) were assessed on day 3. Original magnification: 20× (main images) and 40× (insets); scale bar, 10 μm.

(D) Microglia were purified ex vivo from 5XFAD mice and cultured in 0.1% LCM with or without CSF-1R blocking antibody AFS98. Viability was determined on day 5.

(E and F) Apoptosis of plaque-associated microglia (Iba-1, red) in 5XFAD and Trem2<sup>-/-</sup> 5XFAD mice was determined by TUNEL staining (green). Plaques were identified by X-34 (blue). Representative single-stack images of 5XFAD and Trem2<sup>-/-</sup> 5XFAD microglia (E) and summary of frequencies of TUNEL<sup>+</sup> microglia associated with plaques (F) are shown. Original magnification: 20×; scale bar, 15 μm (E).

(G) Total numbers of live microglia in cortices and hippocampi of 5XFAD, Trem2<sup>-/-</sup> 5XFAD, Trem2<sup>-/-</sup>, and WT mice.

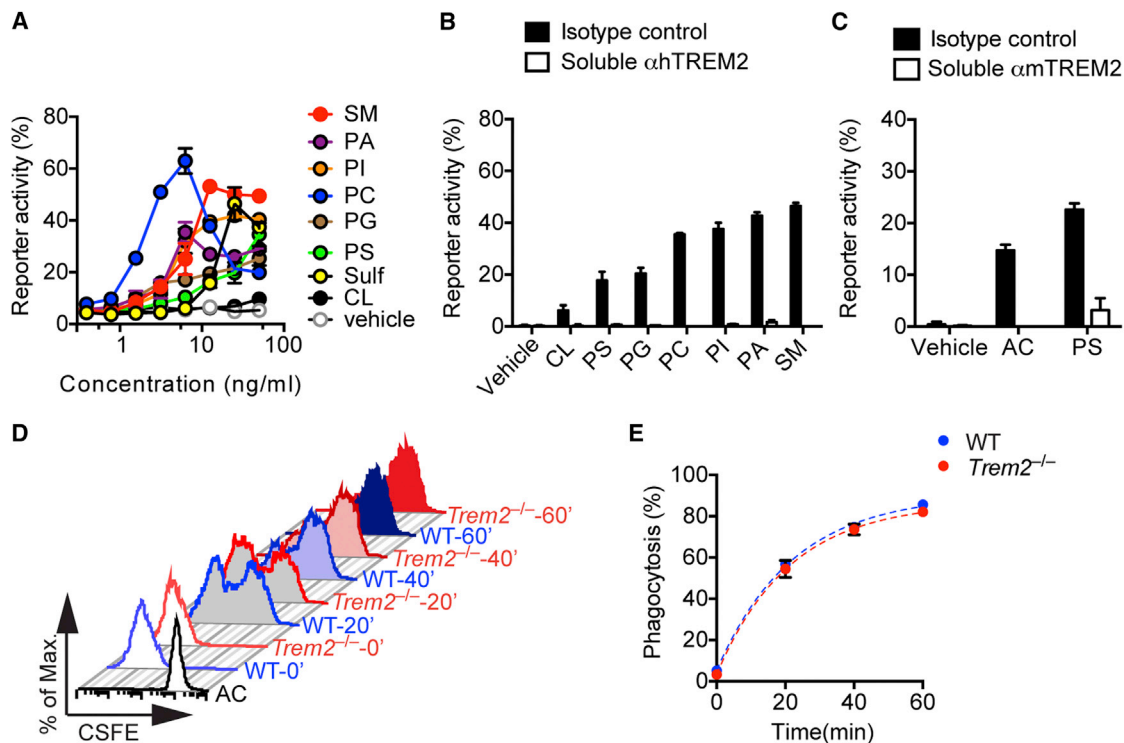
\*\*\*\**p* < 0.0001, two-way ANOVA (A, D, and G), Student's *t* test (F). Data represent a total of three independent experiments (A–D) and a total of five to eight mice per group (E–G). Bars represent mean ± SEM. See also Figure S5 and Movie S4.

might stimulate microglia. These included negatively charged phospholipids, which have been shown to associate with Aβ in lipid membranes (Ahyauch et al., 2012; Nagarathinam et al., 2013); membrane phospholipids, such as phosphatidylserine, which are exposed by damaged neurons and glial cells; and anionic and zwitterionic non-phosphate lipids, such as sulfatides and sphingomyelin, which are released by damaged myelin. We transfected human TREM2 in reporter cells that express GFP under the control of NFAT, such that Ca<sup>2+</sup> mobilization turns on GFP expression when TREM2 is engaged. Incubation of TREM2 reporter cells with many of these lipids activated reporter activity, although to differing extents, with phosphatidylcholine (PC) and sphingomyelin (SM) performing best in these assays (Figures 6A and 6B). Similar results were obtained with a mouse TREM2 reporter (data not shown). Addition of a blocking TREM2 antibody abolished reporter activation by all ligands, demonstrating specificity (Figure 6B). Interestingly, other potential candidates, such as cardiolipin, which is released by damaged mitochondria, did not significantly activate the TREM2 reporter despite its phospholipid structure. This suggests that the ability to engage TREM2 may only partially depend on the presence of negatively charged moieties like phosphoric acid (Figures 6A and 6B). Furthermore, TREM2 reporter activation was not detected with plate-bound synthetic or extracted Aβ (data not shown). In agreement with the ability of phosphatidylserine (PS) to activate

TREM2 reporter cells, apoptotic cells, which expose PS on the cell surface, also activated TREM2 reporter cells (Figure 6C). However, microglia isolated from Trem2<sup>-/-</sup> 5XFAD and 5XFAD mice engulfed apoptotic cells equally well (Figures 6D and 6E). Thus, TREM2 is not directly involved in phagocytosis of apoptotic cells. We conclude that TREM2 is a sensor for several anionic and zwitterionic lipids that are exposed during Aβ deposition as well as during neuronal and glial cell death.

#### R47H Mutation Impairs TREM2 Recognition of Lipid Ligands

What is the impact of the R47H mutation on TREM2 ligand recognition? We generated TREM2 R47H reporter cells and compared their response to identified ligands to that of TREM2 reporter cells. The R47H mutation considerably reduced reporter activation in response to many ligands, including phosphatidic acid (PA), phosphatidylglycerol (PG), PS, phosphatidylinositol (PI), and sulfatides (Figures 7A–7G). The R47H mutation had less impact on SM recognition and very little influence on PC-mediated activation. Importantly, the R47H mutation did not detectably affect cell-surface expression or signaling of TREM2, as assessed by stimulating the R47H reporter cells with a plate-bound anti-TREM2 antibody (Figure 7H). Thus, these data suggest that the R47H reduces the overall capacity of TREM2 to bind anionic ligands.



**Figure 6. TREM2 Is a Receptor for Lipid Patterns Associated with A $\beta$**

(A and B) Human TREM2 reporter cells were stimulated with various phospholipids and anionic and zwitterionic lipids at the indicated concentrations. Reporter activation (GFP expression) was assessed after overnight incubation by flow cytometry. TREM2 reporter cells responding to lipids at various concentrations are shown in (A). Blockade of reporter activation by a soluble anti-hTREM2 mAb is shown in (B). SM, sphingomyelin; PA, phosphatidic acid; PI, phosphatidylinositol; PC, phosphatidylcholine; PG, phosphatidylglycerol; PS, phosphatidylserine; Sulf, sulfatide; and CL, cardiolipin.

(C) mTREM2 reporter cells were cultured with either apoptotic cells (AC) or phosphatidylserine (PS) in the presence of soluble anti-TREM2 mAb or isotype control. (D and E) Adult primary microglia from Trem2<sup>-/-</sup>5XFAD and 5XFAD mice were pulsed with CSFE-labeled AC. (D) Phagocytosis of AC was determined 20, 40, and 60 min post co-culturing by flow cytometry. (E) Summary of AC uptake by WT and Trem2<sup>-/-</sup> microglia.

Data represent a total of three (A–C) and two (D and E) independent experiments.

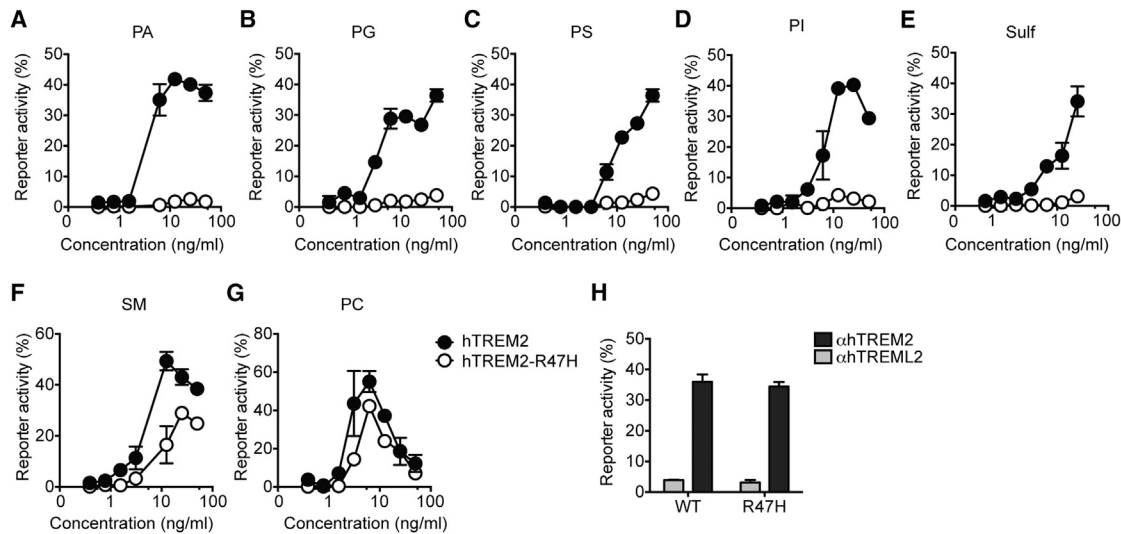
## DISCUSSION

This study showed that TREM2 modulates A $\beta$  accumulation in the 5XFAD mouse model of AD, thereby reducing neuronal damage. The importance of TREM2 in A $\beta$  clearance is underscored by the fact that even the loss of one copy of *Trem2* gene is sufficient to increase A $\beta$  accumulation. TREM2 acts in microglia by supporting A $\beta$ -reactive microgliosis, a process of expansion and activation that leads to microglial clustering around A $\beta$  plaques and subsequent A $\beta$  removal (Ransohoff and Cardona, 2010). In the absence of TREM2, this microgliosis is impaired. In fact, microglia from Trem2<sup>-/-</sup>5XFAD mice are unable to survive, as evidenced by the accumulation of apoptotic microglia around A $\beta$  plaques. Cells involved in TREM2-dependent microgliosis had phenotypic features of brain resident microglia, such as expression of CX3CR1. However, it is possible that monocytes from peripheral blood contribute to microgliosis and that TREM2 supports their survival as well.

Previous studies have shown that CSF-1-CSF-1R signaling is essential for microgliosis in response to A $\beta$  (Chitu and Stanley, 2006; Kaku et al., 2003; Mitrassinovic et al., 2003). Since CSF-1 is rapidly consumed during this process (Stanley and Chitu, 2014), there is probably a limited supply of CSF-1 surrounding

the A $\beta$  plaques. Our results demonstrate that TREM2 provides a signal that is necessary for survival of microglia at low CSF-1 concentrations. We postulate that TREM2 acts as a costimulatory molecule that sustains survival of microglia, which are activated and proliferate in the presence of A $\beta$ . Previous studies of cultured myeloid cells indicate that TREM2 may synergize with CSF-1-CSF-1R signaling to activate the protein tyrosine kinase Syk, which, in turn, activates multiple downstream mediators, such as ERK, PI-3K, and Akt (Zou et al., 2008). In addition, TREM2 may provide survival signals through activation of anti-apoptotic mediators such as  $\beta$ -catenin (Otero et al., 2009) and Mcl-1 (Peng et al., 2010). It is also possible that TREM2 is necessary to support increased microglial metabolism during activation.

Why is TREM2 activated during A $\beta$  accumulation? Previous studies have indicated that TREM2 binds phospholipids, such as PS, and acts as a scavenger receptor for apoptotic cells that might be generated during neuronal damage (Hsieh et al., 2009; Takahashi et al., 2005, 2007). In our study, we demonstrate that TREM2 is a sensor for a broad array of acidic and zwitterionic lipids, which may or may not contain a phosphoric acid moiety. Membranes containing these lipids strongly interact with A $\beta$ , facilitating the formation of fibrillar A $\beta$



**Figure 7. R47H Mutation Attenuates TREM2 Recognition of Lipids**

(A–H) Reporter cells expressing either common allele or R47H variant of human TREM2 were stimulated with various species of lipids or plate-bound anti-hTREM2 mAb. A plate-bound control antibody (anti-hTREM2L2) was used as a negative control. Data represent a total of two independent experiments. Bars represent mean  $\pm$  SEM.

(Ahyayauch et al., 2012; Del Mar Martínez-Senac et al., 1999; Nagarathinam et al., 2013). Moreover, some TREM2 lipidic ligands accumulate on the cell surface of neurons and glial cells damaged by A $\beta$  accumulation, such as PS (Eckert et al., 2005; McLaurin and Chakrabarty, 1996), or are released by damaged myelin, such as SM and sulfatides. In contrast, TREM2 did not directly bind A $\beta$ . Consistent with its ability to bind anionic lipids, the TREM2 extracellular domain is rich in arginine residues that may form salt bridges with polyanions. Remarkably, we found that the R47H mutation associated with AD affected the binding of multiple lipid ligands, although to differing extents. Most likely, the R47H mutation is sufficient to considerably reduce the binding affinity of TREM2 extracellular domain for most anionic lipids. Structural studies will be essential to validate this model.

Our findings demonstrated that TREM2 functions as a microglial sensor that is alerted by damage-induced molecules that share a common lipidic backbone and an anionic group. In contrast with previous reports (Hsieh et al., 2009; Takahashi et al., 2005, 2007), we found that the engagement of TREM2 does not directly mediate phagocytosis of apoptotic cells. However, TREM2 signaling may indirectly support phagocytosis by promoting survival of activated microglia. It has been shown that individuals homozygous for rare mutations that impair expression of either TREM2 or DAP12 develop lethal forms of progressive, early-onset dementia such as Nasu-Hakola disease (NHD) and frontotemporal dementia (Guerreiro et al., 2013a, 2013c; Kleinberger et al., 2014; Paloneva et al., 2002). Although the pathology of these forms of dementia differs from that of AD and often involves demyelination, our study suggests that TREM2 may be required for microglia to sense glycolipids such as SM and sulfatides that are exposed on damaged myelin sheaths; thus, TREM2 binding to these glycolipids may trigger the microglial response to damaged myelin, which is necessary to clear myelin residues and produce trophic factors that induce

repair and remyelination. While the R47H mutation associated with AD did not entirely abolish ligand binding, mutations associated with Nasu-Hakola disease result in a complete lack of TREM2 expression (Kleinberger et al., 2014), which may explain the distinct pathology and more dramatic clinical course of this disease.

## EXPERIMENTAL PROCEDURES

### Mice

*Trem2*<sup>−/−</sup> mice were generated as previously described. 5XFAD mice were purchased from the Jackson Laboratory (MMRRC) and crossed to *Trem2*<sup>−/−</sup> mice to generate *Trem2*<sup>+/−</sup>5XFAD and *Trem2*<sup>−/−</sup>5XFAD mice. All mice were bred and housed in the same animal facility. *Trem2*<sup>−/−</sup>Cx3cr1<sup>+/GFP</sup>APPPS1-21 mice were generated in a similar manner, as previously described (Ulrich et al., 2014). All animal studies were approved by the Washington University Animal Studies Committee.

### Preparation of Brain Samples

For histological analysis 5XFAD mice, APPPS1-21 and transgene negative controls were anesthetized with ketamine and perfused with ice-cold PBS. Right-brain hemispheres were fixed in 4% PFA overnight and placed in 30% sucrose before freezing and cutting on a freezing sliding microtome. Serial 40- $\mu$ m coronal sections of the brain were collected from the rostral anterior commissure to caudal hippocampus as landmarks. For biochemical and mRNA expression analysis, cortices and hippocampi of the left-brain hemispheres were carefully dissected out and flash frozen in liquid nitrogen.

### Immunohistochemistry and Microscopy

For detailed procedures, see Extended Experimental Procedures.

### Gene Expression Analysis

For frozen brain tissues, RNA was extracted using a RNeasy mini kit according to manufacture protocol (QIAGEN). Microglia were fluorescence-activated cell-sorted (FACS) directly into RLT-plus lysis buffer, and RNA extraction was performed using a RNeasy micro kit according to manufacture protocol (QIAGEN). Primers for qPCR analysis are provided in Table S1. For detailed procedure on microarray analysis, see Extended Experimental Procedures.

## ELISA

A $\beta$  levels were assessed using sandwich ELISAs as described (Kim et al., 2009). For detailed procedure, see [Extended Experimental Procedures](#).

## Ex Vivo Microglia Cultures

Primary adult microglia culture was generated as previously described (Butovsky et al., 2014). Briefly, purified adult microglia were cultured in the presence of 15% LCM media (Otero et al., 2009) and 10 ng/ml human TGF- $\beta$ 1 (Pepro-Tech) for 7 days before experiments. For details on in vitro assays performed, see [Extended Experimental Procedures](#).

## Reporter Assay

2B4 GFP-NFAT reporter T cells were stably transfected with murine or human TREM2 cDNAs. Cells were cultured with apoptotic thymocytes in round-bottom 96-well plates or plated onto high-absorbance flat-bottom plate coated with various lipids at indicated concentration. Reporter cells were assessed after overnight incubation. Reporter activity (%) is defined as %GFP<sup>+</sup> cells subtracted from background (vehicle controls).

## Statistics

Data in figures are presented as mean  $\pm$  SEM. All statistical analysis was performed using Prism (GraphPad). Statistical analysis to compare the mean values for multiple groups was performed using a one-way or two-way ANOVA with correction for multiple comparisons. Comparison of two groups was performed using a two-tailed unpaired t test (Mann-Whitney). Values were accepted as significant if  $p \leq 0.05$ .

## ACCESSION NUMBERS

All data have been deposited at GEO (GSE65067).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, one table, and four movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.049>.

## AUTHOR CONTRIBUTIONS

Y.W., J.D.U., M.L.R., S.S., and B.H.Z. performed two-photon and confocal imaging analyses; Y.W., B.H.Z., and G.M.K. performed computational analyses; K.M., K.L.Y., and J.R.C. analyzed A $\beta$  deposition; Y.W. performed ex vivo cell-culture experiments; M. Cella and Y.W. generated TREM2 reporter cells and performed reporter assays; and S.G. generated and maintained *Trem2*<sup>+/-</sup> and *Trem2*<sup>-/-</sup>5XFAD mice. D.M.H. supervised research on APPPS1-21 mice and provided critical comments. M. Colonna supervised the entire project; M. Colonna and Y.W. wrote the manuscript.

## ACKNOWLEDGMENTS

We thank the Genome Technology Access Center (Washington University School of Medicine) for assistance with microarray analysis. We thank S. Raju (Washington University School of Medicine) for assistance with live imaging. We thank R.B. DeMattos, J.D. Sedgewick, and A.P. Martin (Eli Lilly and Company) for helpful suggestions and critical comments. Y.W. is supported by the Lilly Innovation Fellowship Award. B.H.Z. is supported by DDRCC grant P30 DK52574. M. Cella is supported by grant RG4687A1/1 from National Multiple Sclerosis Society. M. Colonna is supported by the Knight Alzheimer's Disease Research Center pilot grant P50 AG005681-30 and the Cure Alzheimer's Fund. Y.W. is an employee of Eli Lilly and Company (Indianapolis, IN).

Received: September 29, 2014

Revised: December 11, 2014

Accepted: January 26, 2015

Published: March 12, 2015

## REFERENCES

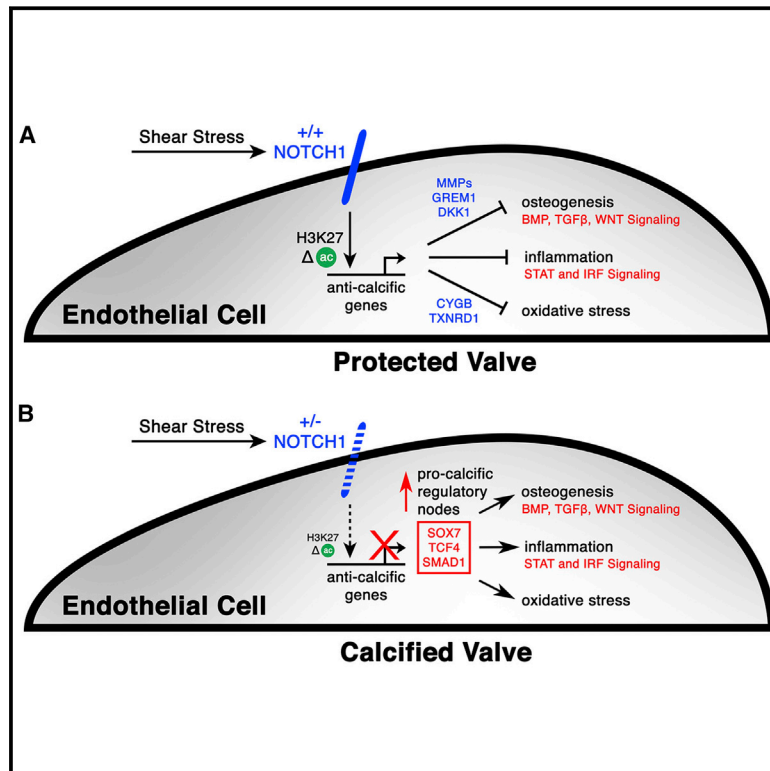
- Ahyayauch, H., Raab, M., Busto, J.V., Andracka, N., Arrondo, J.L., Masserini, M., Tvaroska, I., and Goñi, F.M. (2012). Binding of  $\beta$ -amyloid (1-42) peptide to negatively charged phospholipid membranes in the liquid-ordered state: modeling and experimental studies. *Biophys. J.* 103, 453–463.
- Bouchon, A., Hernández-Munain, C., Cella, M., and Colonna, M. (2001). A DAP12-mediated pathway regulates expression of CC chemokine receptor 7 and maturation of human dendritic cells. *J. Exp. Med.* 194, 1111–1122.
- Butovsky, O., Jedrychowski, M.P., Moore, C.S., Cialic, R., Lanser, A.J., Gabrieli, G., Koeglsperger, T., Dake, B., Wu, P.M., Doykan, C.E., et al. (2014). Identification of a unique TGF- $\beta$ -dependent molecular and functional signature in microglia. *Nat. Neurosci.* 17, 131–143.
- Cannon, J.P., O'Driscoll, M., and Litman, G.W. (2012). Specific lipid recognition is a general feature of CD300 and TREM molecules. *Immunogenetics* 64, 39–47.
- Chitu, V., and Stanley, E.R. (2006). Colony-stimulating factor-1 in immunity and inflammation. *Curr. Opin. Immunol.* 18, 39–48.
- D'Andrea, M.R., Cole, G.M., and Ard, M.D. (2004). The microglial phagocytic role with specific plaque types in the Alzheimer disease brain. *Neurobiol. Aging* 25, 675–683.
- Daws, M.R., Lanier, L.L., Seaman, W.E., and Ryan, J.C. (2001). Cloning and characterization of a novel mouse myeloid DAP12-associated receptor family. *Eur. J. Immunol.* 31, 783–791.
- Daws, M.R., Sullam, P.M., Niemi, E.C., Chen, T.T., Tchao, N.K., and Seaman, W.E. (2003). Pattern recognition by TREM-2: binding of anionic ligands. *J. Immunol.* 171, 594–599.
- Del Mar Martínez-Senac, M., Villalain, J., and Gómez-Fernández, J.C. (1999). Structure of the Alzheimer beta-amyloid peptide (25-35) and its interaction with negatively charged phospholipid vesicles. *Eur. J. Biochem.* 265, 744–753.
- Dickson, D.W. (1999). Microglia in Alzheimer's disease and transgenic models. How close the fit? *Am. J. Pathol.* 154, 1627–1631.
- Eckert, G.P., Wood, W.G., and Müller, W.E. (2005). Membrane disordering effects of beta-amyloid peptides. *Subcell. Biochem.* 38, 319–337.
- Eimer, W.A., and Vassar, R. (2013). Neuron loss in the 5XFAD mouse model of Alzheimer's disease correlates with intraneuronal A $\beta$ 42 accumulation and Caspase-3 activation. *Mol. Neurodegener.* 8, 2.
- El Khoury, J., Toft, M., Hickman, S.E., Means, T.K., Terada, K., Geula, C., and Luster, A.D. (2007). Ccr2 deficiency impairs microglial accumulation and accelerates progression of Alzheimer-like disease. *Nat. Med.* 13, 432–438.
- Ford, J.W., and McVicar, D.W. (2009). TREM and TREM-like receptors in inflammation and disease. *Curr. Opin. Immunol.* 21, 38–46.
- Frautschy, S.A., Yang, F., Irrizarry, M., Hyman, B., Saido, T.C., Hsiao, K., and Cole, G.M. (1998). Microglial response to amyloid plaques in APPsw transgenic mice. *Am. J. Pathol.* 152, 307–317.
- Grathwohl, S.A., Kälin, R.E., Bolmont, T., Prokop, S., Winkemann, G., Kaeser, S.A., Odenthal, J., Radde, R., Eldh, T., Gandy, S., et al. (2009). Formation and maintenance of Alzheimer's disease beta-amyloid plaques in the absence of microglia. *Nat. Neurosci.* 12, 1361–1363.
- Guerreiro, R., Bilgic, B., Guven, G., Bras, J., Rohrer, J., Lohmann, E., Hanagasi, H., Gurvit, H., and Emre, M. (2013a). Novel compound heterozygous mutation in TREM2 found in a Turkish frontotemporal dementia-like family. *Neurobiol. Aging* 34, 2890, e2891–2895.
- Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S., Younkin, S., et al.; Alzheimer Genetic Analysis Group (2013b). TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* 368, 117–127.
- Guerreiro, R.J., Lohmann, E., Brás, J.M., Gibbs, J.R., Rohrer, J.D., Gurunlian, N., Dursun, B., Bilgic, B., Hanagasi, H., Gurvit, H., et al. (2013c). Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA Neurol.* 70, 78–84.
- Hamerman, J.A., Jarjoura, J.R., Humphrey, M.B., Nakamura, M.C., Seaman, W.E., and Lanier, L.L. (2006). Cutting edge: inhibition of TLR and FcR



- responses in macrophages by triggering receptor expressed on myeloid cells (TREM)-2 and DAP12. *J. Immunol.* 177, 2051–2055.
- Hickman, S.E., Allison, E.K., and El Khoury, J. (2008). Microglial dysfunction and defective beta-amyloid clearance pathways in aging Alzheimer's disease mice. *J. Neurosci.* 28, 8354–8360.
- Hsieh, C.L., Koike, M., Spusta, S.C., Niemi, E.C., Yenari, M., Nakamura, M.C., and Seaman, W.E. (2009). A role for TREM2 ligands in the phagocytosis of apoptotic neuronal cells by microglia. *J. Neurochem.* 109, 1144–1156.
- Huang, Y., and Mucke, L. (2012). Alzheimer mechanisms and therapeutic strategies. *Cell* 148, 1204–1222.
- Jiang, T., Tan, L., Zhu, X.C., Zhang, Q.Q., Cao, L., Tan, M.S., Gu, L.Z., Wang, H.F., Ding, Z.Z., Zhang, Y.D., et al. (2014). Upregulation of TREM2 Ameliorates Neuropathology and Rescues Spatial Cognitive Impairment in a Transgenic Mouse Model of Alzheimer's Disease. *Neuropsychopharmacology*.
- Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., et al. (2013). Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* 368, 107–116.
- Kaku, M., Tsutsui, K., Motokawa, M., Kawata, T., Fujita, T., Kohno, S., Tohma, Y., Ohtani, J., Tenjoh, K., and Tanne, K. (2003). Amyloid beta protein deposition and neuron loss in osteopetrotic (op/op) mice. *Brain Res. Brain Res. Protoc.* 12, 104–108.
- Kiialainen, A., Hovanes, K., Paloneva, J., Kopra, O., and Peltonen, L. (2005). Dap12 and Trem2, molecules involved in innate immunity and neurodegeneration, are co-expressed in the CNS. *Neurobiol. Dis.* 18, 314–322.
- Kim, J., Basak, J.M., and Holtzman, D.M. (2009). The role of apolipoprotein E in Alzheimer's disease. *Neuron* 63, 287–303.
- Kleinberger, G., Yamanishi, Y., Suarez-Calvet, M., Czirr, E., Lohmann, E., Cuyvers, E., Struyfs, H., Pettkus, N., Wenninger-Weinzierl, A., Mazaheri, F., et al. (2014). TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Sci Transl Med* 6, 243ra286.
- Malm, T.M., Koistinaho, M., Pärepa, M., Vatanen, T., Ooka, A., Karlsson, S., and Koistinaho, J. (2005). Bone-marrow-derived cells contribute to the recruitment of microglial cells in response to beta-amyloid deposition in APP/PS1 double transgenic Alzheimer mice. *Neurobiol. Dis.* 18, 134–142.
- McGeer, P.L., Itagaki, S., Tago, H., and McGeer, E.G. (1987). Reactive microglia in patients with senile dementia of the Alzheimer type are positive for the histocompatibility glycoprotein HLA-DR. *Neurosci. Lett.* 79, 195–200.
- McLaurin, J., and Chakrabarty, A. (1996). Membrane disruption by Alzheimer beta-amyloid peptides mediated through specific binding to either phospholipids or gangliosides. Implications for neurotoxicity. *J. Biol. Chem.* 271, 26482–26489.
- Mildner, A., Schlevogt, B., Kierdorf, K., Böttcher, C., Emy, D., Kummer, M.P., Quinn, M., Brück, W., Bechmann, I., Heneka, M.T., et al. (2011). Distinct and non-redundant roles of microglia and myeloid subsets in mouse models of Alzheimer's disease. *J. Neurosci.* 31, 11159–11171.
- Mitrasinovic, O.M., Vincent, V.A., Simsek, D., and Murphy, G.M., Jr. (2003). Macrophage colony stimulating factor promotes phagocytosis by murine microglia. *Neurosci. Lett.* 344, 185–188.
- Nagarathinam, A., Höflinger, P., Bühler, A., Schäfer, C., McGovern, G., Jeffrey, M., Staufenbiel, M., Jucker, M., and Baumann, F. (2013). Membrane-anchored Abeta accelerates amyloid formation and exacerbates amyloid-associated toxicity in mice. *J. Neurosci.* 33, 19284–19294.
- Oakley, H., Cole, S.L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Van Eldik, L., et al. (2006). Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J. Neurosci.* 26, 10129–10140.
- Otero, K., Turnbull, I.R., Poliani, P.L., Vermi, W., Cerutti, E., Aoshi, T., Tassi, I., Takai, T., Stanley, S.L., Miller, M., et al. (2009). Macrophage colony-stimulating factor induces the proliferation and survival of macrophages via a pathway involving DAP12 and beta-catenin. *Nat. Immunol.* 10, 734–743.
- Otero, K., Shinohara, M., Zhao, H., Cella, M., Gilfillan, S., Colucci, A., Faccio, R., Ross, F.P., Teitelbaum, S.L., Takayanagi, H., and Colonna, M. (2012). TREM2 and  $\beta$ -catenin regulate bone homeostasis by controlling the rate of osteoclastogenesis. *J. Immunol.* 188, 2612–2621.
- Paloneva, J., Manninen, T., Christman, G., Hovanes, K., Mandelin, J., Adolfsson, R., Bianchin, M., Bird, T., Miranda, R., Salmaggi, A., et al. (2002). Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. *Am. J. Hum. Genet.* 71, 656–662.
- Peng, Q., Malhotra, S., Torchia, J.A., Kerr, W.G., Coggeshall, K.M., and Humphrey, M.B. (2010). TREM2- and DAP12-dependent activation of PI3K requires DAP10 and is inhibited by SHIP1. *Sci. Signal.* 3, ra38.
- Perlmutter, L.S., Barron, E., and Chui, H.C. (1990). Morphologic association between microglia and senile plaque amyloid in Alzheimer's disease. *Neurosci. Lett.* 119, 32–36.
- Ransohoff, R.M., and Cardona, A.E. (2010). The myeloid cells of the central nervous system parenchyma. *Nature* 468, 253–262.
- Schmid, C.D., Sautkulis, L.N., Danielson, P.E., Cooper, J., Hasel, K.W., Hilbush, B.S., Sutcliffe, J.G., and Carson, M.J. (2002). Heterogeneous expression of the triggering receptor expressed on myeloid cells-2 on adult murine microglia. *J. Neurochem.* 83, 1309–1320.
- Sessa, G., Podini, P., Mariani, M., Meroni, A., Spreafico, R., Sinigaglia, F., Colonna, M., Panina, P., and Meldolesi, J. (2004). Distribution and signaling of TREM2/DAP12, the receptor system mutated in human polycystic lipomembraneous osteodysplasia with sclerosing leukoencephalopathy dementia. *Eur. J. Neurosci.* 20, 2617–2628.
- Simard, A.R., Soulet, D., Gowing, G., Julien, J.P., and Rivest, S. (2006). Bone marrow-derived microglia play a critical role in restricting senile plaque formation in Alzheimer's disease. *Neuron* 49, 489–502.
- Stalder, M., Phinney, A., Probst, A., Sommer, B., Staufenbiel, M., and Jucker, M. (1999). Association of microglia with amyloid plaques in brains of APP23 transgenic mice. *Am. J. Pathol.* 154, 1673–1684.
- Stalder, A.K., Ermini, F., Bondolfi, L., Krenger, W., Burbach, G.J., Deller, T., Coomaraswamy, J., Staufenbiel, M., Landmann, R., and Jucker, M. (2005). Invasion of hematopoietic cells into the brain of amyloid precursor protein transgenic mice. *J. Neurosci.* 25, 11125–11132.
- Stanley, E.R., and Chitu, V. (2014). CSF-1 receptor signaling in myeloid cells. *Cold Spring Harb. Perspect. Biol.* 6 <http://dx.doi.org/10.1101/cshperspect.a021857>.
- Streit, W.J., and Xue, Q.S. (2009). Life and death of microglia. *J. Neuroimmune Pharmacol.* 4, 371–379.
- Streit, W.J., Sammons, N.W., Kuhns, A.J., and Sparks, D.L. (2004). Dystrophic microglia in the aging human brain. *Glia* 45, 208–212.
- Takahashi, K., Rochford, C.D., and Neumann, H. (2005). Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *J. Exp. Med.* 201, 647–657.
- Takahashi, K., Prinz, M., Stagi, M., Chechneva, O., and Neumann, H. (2007). TREM2-transduced myeloid precursors mediate nervous tissue debris clearance and facilitate recovery in an animal model of multiple sclerosis. *PLoS Med.* 4, e124.
- Tanzi, R.E. (2013). A brief history of Alzheimer's disease gene discovery. *J. Alzheimers Dis.* 33 (Suppl 1), S5–S13.
- Thrash, J.C., Torbett, B.E., and Carson, M.J. (2009). Developmental regulation of TREM2 and DAP12 expression in the murine CNS: implications for Nasu-Hakola disease. *Neurochem. Res.* 34, 38–45.
- Turnbull, I.R., Gilfillan, S., Cella, M., Aoshi, T., Miller, M., Piccio, L., Hernandez, M., and Colonna, M. (2006). Cutting edge: TREM-2 attenuates macrophage activation. *J. Immunol.* 177, 3520–3524.
- Ulrich, J.D., Finn, M.B., Wang, Y., Shen, A., Mahan, T.E., Jiang, H., Stewart, F.R., Piccio, L., Colonna, M., and Holtzman, D.M. (2014). Altered microglial response to A $\beta$  plaques in APPPS1-21 mice heterozygous for TREM2. *Mol. Neurodegener.* 9, 20.
- Zou, W., Reeve, J.L., Liu, Y., Teitelbaum, S.L., and Ross, F.P. (2008). DAP12 couples c-Fms activation to the osteoclast cytoskeleton by recruitment of Syk. *Mol. Cell* 31, 422–431.

# Human Disease Modeling Reveals Integrated Transcriptional and Epigenetic Mechanisms of NOTCH1 Haploinsufficiency

## Graphical Abstract



## Authors

Christina V. Theodoris, Molong Li, ..., Benoit G. Bruneau, Deepak Srivastava

## Correspondence

dsrivastava@gladstone.ucsf.edu

## In Brief

Heterozygous mutations in *NOTCH1* that cause aortic valve calcification disrupt the transcriptional and epigenetic response to hemodynamic shear stress, resulting in derepression of latent pro-osteogenic and -inflammatory gene networks.

## Highlights

- Shear stress protects human endothelial cells by suppressing osteogenesis and inflammation
- NOTCH1 (N1) mediates the normal anti-calcific response induced by shear stress
- *N1* haploinsufficiency leads to differential H3K27ac at N1-bound enhancers
- Intervention at key regulatory nodes can alleviate N1-dependent network dysregulation



# Human Disease Modeling Reveals Integrated Transcriptional and Epigenetic Mechanisms of NOTCH1 Haploinsufficiency

Christina V. Theodoris,<sup>1,2,3</sup> Molong Li,<sup>1,2</sup> Mark P. White,<sup>1,2</sup> Lei Liu,<sup>1,2</sup> Daniel He,<sup>1,2</sup> Katherine S. Pollard,<sup>1,4,8</sup> Benoit G. Bruneau,<sup>1,2,3,5,6,8</sup> and Deepak Srivastava<sup>1,2,3,5,7,8,\*</sup>

<sup>1</sup>Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA

<sup>2</sup>Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA 94158, USA

<sup>3</sup>Program in Developmental and Stem Cell Biology

<sup>4</sup>Department of Epidemiology and Biostatistics

<sup>5</sup>Department of Pediatrics

<sup>6</sup>Cardiovascular Research Institute

<sup>7</sup>Department of Biochemistry and Biophysics

University of California, San Francisco, San Francisco, CA 94158, USA

<sup>8</sup>Co-senior author

\*Correspondence: [dsrivastava@gladstone.ucsf.edu](mailto:dsrivastava@gladstone.ucsf.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.035>

## SUMMARY

The mechanisms by which transcription factor haploinsufficiency alters the epigenetic and transcriptional landscape in human cells to cause disease are unknown. Here, we utilized human induced pluripotent stem cell (iPSC)-derived endothelial cells (ECs) to show that heterozygous nonsense mutations in *NOTCH1* that cause aortic valve calcification disrupt the epigenetic architecture, resulting in derepression of latent pro-osteogenic and -inflammatory gene networks. Hemodynamic shear stress, which protects valves from calcification in vivo, activated anti-osteogenic and anti-inflammatory networks in *NOTCH1*<sup>+/+</sup>, but not *NOTCH1*<sup>+/-</sup>, iPSC-derived ECs. *NOTCH1* haploinsufficiency altered H3K27ac at *NOTCH1*-bound enhancers, dysregulating downstream transcription of more than 1,000 genes involved in osteogenesis, inflammation, and oxidative stress. Computational predictions of the disrupted *NOTCH1*-dependent gene network revealed regulatory nodes that, when modulated, restored the network toward the *NOTCH1*<sup>+/+</sup> state. Our results highlight how alterations in transcription factor dosage affect gene networks leading to human disease and reveal nodes for potential therapeutic intervention.

## INTRODUCTION

Human disease is often caused by genetic variants that quantitatively affect dosage of the encoded gene product, particularly those involving major regulatory factors. The use of induced pluripotent stem cells (iPSCs) has facilitated the understanding of many human diseases, but it remains unclear how reduction

in dosage of transcriptional regulators selectively affects the transcription of target genes, alters the epigenetic landscape, and perturbs gene networks resulting in disease. The ability to model haploinsufficiency of a transcription factor (TF) in human iPSCs combined with integration of broad “-omic” data may reveal mechanisms underlying dose sensitivity of regulatory proteins and novel targets for intervention.

We previously reported two families with heterozygous nonsense mutations in the membrane-bound TF, *NOTCH1* (N1), which led to a congenital defect of the aortic valve known as bicuspid aortic valve (BAV) and severe aortic valve calcification in adults (Garg et al., 2005). Calcific aortic valve disease (CAVD) is the third leading cause of adult heart disease and is responsible for more than 100,000 valve transplants annually in the United States alone (Garg et al., 2005). BAV, which occurs in 1%–2% of the population and involves the formation of two valve leaflets rather than the normal three leaflets, is a major risk factor for early valve calcification, although the mechanism for the calcification is unknown (Go et al., 2014). Recent studies identified N1 mutations in additional familial cases of BAV and CAVD, as well as ~4% of sporadic cases, underscoring the importance of N1 in this disease (Foffa et al., 2013; Mohamed et al., 2006).

Hemodynamic shear stress protects against aortic valve calcification in adults, similar to shear-induced protection against atherosclerosis and vascular calcification. Accordingly, the first region of the valve to calcify is the aortic side, which experiences less laminar shear stress than the ventricular side (Weinberg et al., 2010). Shear stress activates signaling through the N1 transmembrane receptor in endothelial cells (ECs) in vitro, and *NOTCH* signaling in vivo is greater on the ventricular side of the aortic valve (Combs and Yutzey, 2009; Masumura et al., 2009). Furthermore, in mice, EC-specific deletion of the Notch ligand Jagged1 leads to valve malformations and aortic valve calcification (Hofmann et al., 2012). These findings suggest that N1 signaling in the endothelium is uniquely positioned to mediate the anti-calcific response to shear stress within the valve.

Here, we utilized human iPSC-derived ECs to show that heterozygous nonsense mutations in *N1* disrupt the epigenetic architecture, resulting in derepression of latent pro-osteogenic and -inflammatory gene networks. Hemodynamic shear stress activated anti-osteogenic and anti-inflammatory networks in *N1*<sup>+/+</sup>, but not *N1*<sup>+/-</sup>, iPSC-derived ECs. *N1* haploinsufficiency altered H3K27ac at *N1*-bound enhancers, dysregulating downstream transcription of more than 1,000 genes involved in osteogenesis, inflammation, and oxidative stress. Computational predictions of the disrupted *N1*-dependent gene network revealed regulatory nodes that, when modulated, restored the network toward the wild-type (WT) state. Our results highlight how alterations in TF dosage affect gene networks leading to human disease and reveal nodes for potential therapeutic intervention.

## RESULTS

### Transcriptional States in EC Differentiation and Response to Shear Stress

To investigate the consequences of *N1* heterozygosity in ECs, we first needed to describe the normal transcriptional and epigenetic state of human ECs during differentiation and under static and fluid shear stress conditions. We therefore differentiated two human embryonic stem cell (ESC) lines (H7 and H9) and three human iPSC lines into ECs using a protocol previously developed in our lab (Figure 1A) (White et al., 2013). We collected cells at key stages of EC differentiation: undifferentiated pluripotent cells, mesodermal precursors (MesoPs), EC precursors (ECPs), and ECs that we exposed to either static or laminar shear stress conditions to model the effects of hemodynamic shear stress on the ventricular side of the aortic valve (Figure 1A). We only conducted experiments on ECPs and ECs that were 70%–100% pure for their respective markers by fluorescence-activated cell sorting (FACS) (Figures S1A and S1B).

We first identified the unique signature of key stages of EC differentiation using RNA sequencing (RNA-seq) data from each aforementioned cell population (Figure 1B and Tables S1 and S2). As expected, genes related to cell division and stem cell maintenance defined pluripotent cells, whereas genes involved in WNT, HEDGEHOG, and BMP signaling were enriched in MesoPs. By the ECP stage, genes involved in angiogenesis and MAPK signaling were upregulated, indicating the commencement of EC specification. NOTCH signaling was a unique feature of the final shear-responsive EC stage. ECs also showed upregulation of matrix metalloproteinases (MMPs), which are involved in degrading extracellular matrix (ECM) (Vu and Werb, 2000).

Genes upregulated in shear stress conditions were involved in antagonizing pro-osteogenic BMP and TGF $\beta$  signaling pathways and included TFs such as SMAD6 and SMAD7 and secreted factors such as GREM1 (Figure 1B) (Bragdon et al., 2011). The secreted factors may provide a mechanism for ECs exposed to shear stress to prevent the calcification of underlying valve interstitial cells (VICs). Additionally, shear-stress-upregulated genes involved factors that increase the resistance to oxidative damage, including *NQO1* and *TXNRD1* (Gorini et al., 2013). Using a random forest machine learning approach to

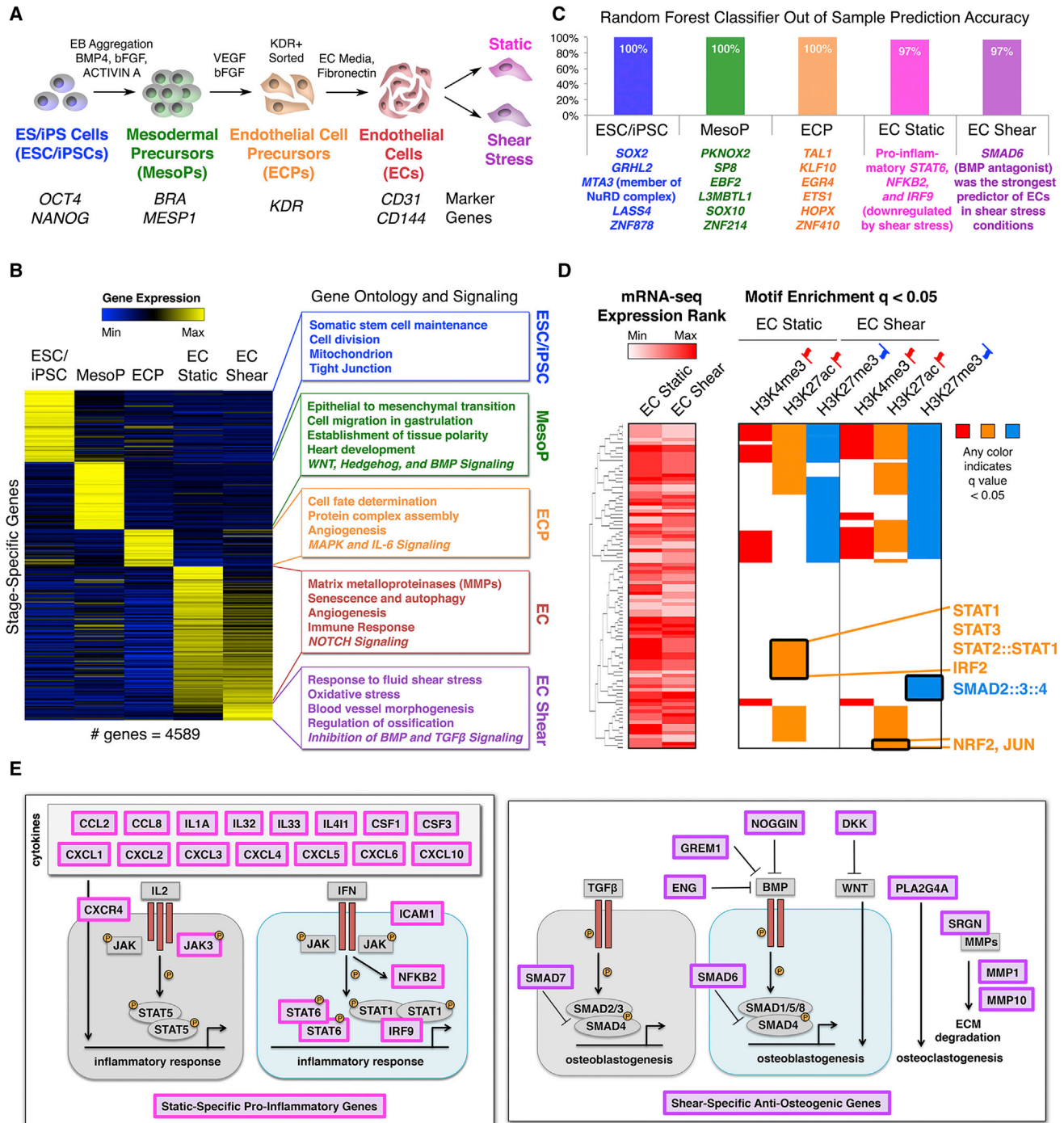
identify stage-predictive TFs from the RNA-seq data, we found that static conditions were predicted by expression of genes encoding inflammatory proteins, including STAT6, NFKB2, and IRF9, all of which were downregulated by shear stress, while the TF most predictive of shear stress conditions was SMAD6, which inhibits BMP signaling (Figures 1C and S1C) (Hervas-Stubbs et al., 2011).

To investigate whether the anti-inflammatory and anti-osteogenic effects of shear stress were mediated by changes in genome occupancy of these key TFs, we tested the distribution of their putative occupancy sites based on motif analysis within active enhancers or repressed regions as identified by H3K27ac and H3K27me3 chromatin immunoprecipitation sequencing (ChIP-seq), respectively, in human iPSC-derived ECs (Figure 1D). In static conditions, we found a unique overrepresentation of pro-inflammatory STAT and IRF motifs in H3K27ac-marked enhancers. By contrast, H3K27ac enhancers in the shear stress condition showed a unique overrepresentation of anti-inflammatory and anti-oxidant NRF2 motifs and TGF $\beta$ -inhibitory JUN motifs (Dennler et al., 2000; Gorini et al., 2013). H3K27me3-marked repressive regions in shear stress conditions showed a unique overrepresentation of SMAD2/3/4 motifs, suggesting repression of pro-osteogenic TGF $\beta$  signaling in shear stress conditions. Thus, shear stress may protect against calcification by antagonizing pro-osteogenic BMP and TGF $\beta$  signaling pathways and repressing pro-inflammatory STAT and IRF signaling pathways at the transcriptional and epigenetic level (Figure 1E).

### Dynamic Chromatin States Correlated with Distinct Transcriptional Patterns

To understand the epigenetic changes occurring near the transcriptional start sites (TSSs) of dynamically expressed genes, we performed ChIP-seq for the H3K4me3 active promoter mark, H3K27ac active enhancer mark, H3K4me1 poised/active mark, and H3K27me3 repressive mark across EC differentiation (Rada-Iglesias et al., 2011; Wamstad et al., 2012) (Figures 2A and 2B and Table S3). Dynamic changes in histone modifications at promoters fell into distinct clusters. To test whether dynamic histone modifications correlated with distinct transcriptional patterns to distinguish functional groups of co-expressed genes, we compiled genes shared between each chromatin and expression cluster and determined statistical enrichment (Figure 2C). Most expression clusters correlated with multiple chromatin clusters. For example, expression cluster F contained genes highly transcribed at the MesoP stage, which were enriched in chromatin clusters 3–8 and 14–15. Previous studies have shown that genes important for early development often correlate with a chromatin modification pattern similar to that of chromatin cluster 8, which includes a high level of repressive H3K27me3 at the pluripotent stage that is then relieved to allow expression during early development (Rada-Iglesias et al., 2011). Indeed, we found that there was a significant enrichment in cluster F of genes annotated as developmental within chromatin cluster 8 compared to genes annotated as non-developmental (Figure S1D). This indicated that chromatin cluster 8 identified a specific functional subset of genes in expression cluster F.





**Figure 1. Transcriptional Mechanisms in EC Differentiation and Response to Shear Stress**

(A) Stages of EC differentiation analyzed.

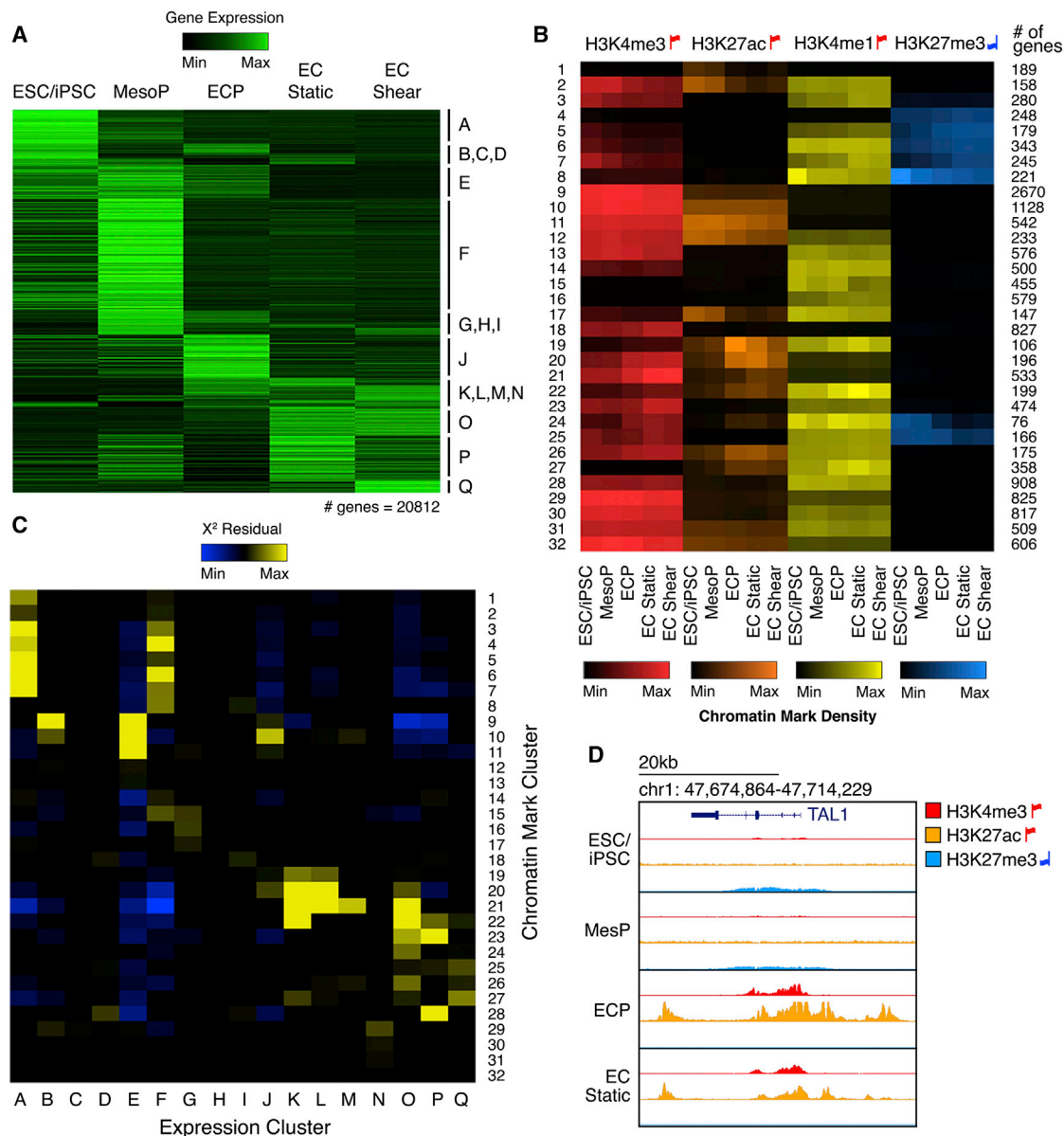
(B) Unique signature of EC differentiation stages by RNA-seq. Stage-unique genes were expressed most highly at the given stage and significantly upregulated relative to immediately preceding or following stages.  $p < 0.05$  by negative binomial test with false discovery rate (FDR) correction.

(C) Top stage-predictive TFs identified by random forest classifier.

(D) Left: expression of TFs whose motifs were tested in the corresponding rows on the right. Right: motif enrichment within activating or repressed signaling pathways. Any color indicates significant motif enrichment ( $q < 0.05$ ) by motifDiverge with FDR correction, whereas white indicates non-significance. Red up flags: activating marks; blue down flags: repressive marks.

(E) Left: diagram of static-specific pro-inflammatory genes (pink). Right: diagram of shear-specific anti-osteogenic genes (violet).

In (B–D):  $n = 5$ . See also Figure S1 and Tables S1 and S2.



**Figure 2. Correlation of Dynamic Chromatin Patterns with Transcriptional Transitions**

(A) Hierarchical clustering of mRNA expression.

(B) Hierarchical clustering of genes based on enrichment of histone modifications within 1 kb of the TSS. Color indicates mean enrichment for each gene cluster. Red up flags: activating marks; blue down flags: repressive marks.

(C) The overlap of genes within expression clusters (horizontal axis) and chromatin clusters (vertical axis). Color represents X<sup>2</sup> residuals (any yellow indicates significant overlap between genes in the corresponding expression and chromatin cluster).

(D) Histone modification enrichment around *TAL1* (Cluster J) during EC differentiation.

In (A–D): n = 5. See also Figure S1 and Tables S3 and S4.

Conversely, some expression clusters correlated strongly with a single chromatin cluster. For example, expression cluster J predominantly clustered with chromatin cluster 10. While genes expressed in J showed very specific upregulation at the ECP stage, chromatin cluster 10 was characterized by stable H3K4me3 and H3K27ac activating marks throughout differentiation. All other expression clusters correlating with this non-

dynamic chromatin cluster included genes upregulated at the ECP stage. This pattern was upheld even when evaluating chromatin marks up to 15 kilobases (kb) from gene TSSs.

The aforementioned random forest approach to discern stage-predictive TFs identified several ECP-predictive TFs that may confer the specificity of cluster J genes, including *SCL1/TAL1*, an important regulator of ECPs and vasculogenesis, in

addition to hematopoiesis (Drake et al., 1997; Liao et al., 1998; Van Handel et al., 2012) (Figure 1C). Unlike other cluster J genes that showed consistently active promoters throughout differentiation, H3K4me3 and H3K27ac marked the *TAL1* promoter most strongly at the ECP stage (Figure 2D), when *TAL1* drives critical fate decisions (Van Handel et al., 2012). Additionally, repressive H3K27me3 marked the *TAL1* promoter at prior stages, suggesting that its repression may prevent premature activation of downstream ECP gene networks. *TAL1* gene expression was elevated in ECPs, but *TAL1* DNA-binding motifs were not enriched in any stage-specific enhancers throughout EC differentiation. Thus, genes expressed at the pivotal ECP specification stage maintain an epigenetic state primed for transcriptional activation throughout differentiation and may rely on the expression of a discrete set of regulators such as *TAL1* to confer temporal specificity. Only 11 other cluster J genes shared *TAL1*'s dynamic chromatin pattern (chromatin cluster 24). These included *GATA2*, a known cofactor of EC regulator *ETV2* (Shi et al., 2014), as well as genes with no previously known role in EC differentiation, such as *C16orf74*, that may represent novel ECP regulators. Together, these ECP-enriched genes and their targets activate genes critical for downstream endothelial development such as *ANGPT2*, a known *TAL1* target gene that is required for postnatal angiogenesis (Deleuze et al., 2012), maintaining a high level of activation in ECs under both static and shear stress conditions (Figure S1E).

We next investigated whether chromatin clusters could discern functional groups of genes involved in the response to shear stress. Two groups of expression clusters contained static or shear-stress-specific gene expression: K–L and P–Q. While K–L was enriched for cytoskeletal genes, P–Q was enriched for genes involved in SMAD signaling and blood vessel development. We focused on expression clusters P and Q to test whether the associated chromatin clusters could distinguish functional gene groups important for the anti-calcific effects of shear stress (Table S4). Gene ontology (GO) term enrichment in individual expression-chromatin cluster intersections showed that static-specific cluster intersections P22 and P25 contained immune process and interleukin signaling factors, respectively. However, when intersected with shear-specific genes, the same chromatin clusters, 22 and 25, were associated with TGF $\beta$  signaling antagonism (including *ENG* and *INHBA*) and bone mineralization (including inhibitory *GREM1*) (Bragdon et al., 2011; Guo et al., 2004). Additional shear-specific cluster intersection Q24 contained activin-binding genes such as *FSTL3*, an inhibitor of TGF $\beta$  signaling (Bragdon et al., 2011). Thus, chromatin clusters were able to distinguish immune process-related genes active in the static condition and antagonism of pro-osteogenic TGF $\beta$  signaling in shear stress conditions.

### Isogenic iPSC-Derived ECs Model N1 Haploinsufficiency

We sought to understand how *N1* heterozygosity perturbs the normal EC gene expression and epigenetic state to cause CAVD. We derived and characterized iPSCs from the fibroblasts of three individuals from two families affected with CAVD due to heterozygous nonsense mutations in *N1* (Figures 3A, S2A–S2E, and S3). Additionally, we derived and characterized iPSCs from a related individual who was *N1*<sup>+/+</sup> and unaffected by CAVD.

As unrelated controls, we used two established ESC lines (H7 and H9) and two previously established iPSC lines.

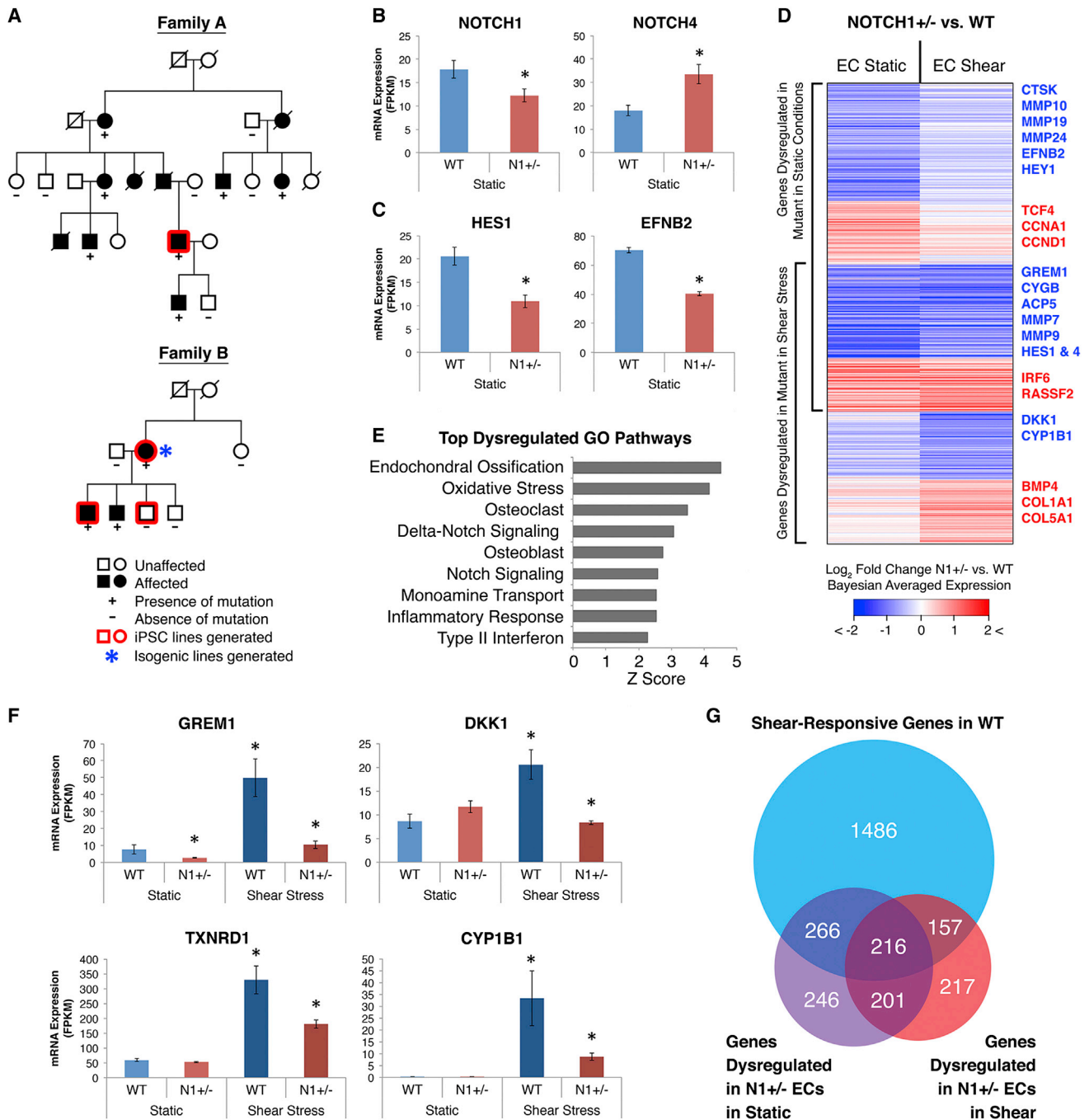
To derive isogenic control lines, we corrected the *N1* mutation using TALENs (Figures 3A and S4A–S4D). We detected no off-target effects by Southern blot for the donor DNA and compared multiple corrected *N1*<sup>+/+</sup> clones to multiple TALEN-targeted but uncorrected *N1*<sup>+/-</sup> clones to control for any effects of the TALEN-targeting process. We differentiated control and mutant iPSCs into ECs and exposed them to either static or shear stress conditions to model the anti-calcific effects of shear stress.

Using isogenic cell lines, we found that *N1* mRNA levels were reduced by 30%–40% in the *N1*<sup>+/-</sup> ECs by RNA-seq (Figure 3B). Additionally, 70%–100% of *N1* mRNA present in *N1*<sup>+/-</sup> ECs was transcribed from the WT allele, suggesting that the mutant mRNA largely undergoes nonsense-mediated decay (Figure S5D). This indicated that the iPSC-derived ECs were effectively modeling a decreased dosage of *N1*. *N1*<sup>+/-</sup> ECs showed increased levels of *NOTCH4* mRNA, which encodes the other major NOTCH protein in ECs and reflects a possible compensatory response. Although *N1*<sup>+/-</sup> ECs did not exhibit altered differentiation capacity, their transcriptome clustered separately from isogenic *N1*<sup>+/+</sup> ECs (Figures S4C and S5A). Furthermore, canonical *N1* targets, including *HES1* and *EFNB2*, were downregulated in *N1*<sup>+/-</sup> ECs, indicating that *N1*<sup>+/-</sup> ECs were haploinsufficient for *N1* activity, at least at some targets (Figure 3C).

### Gene Networks Dysregulated due to N1 Haploinsufficiency

RNA-seq of iPSC-derived ECs in static conditions revealed that 929 mRNAs were dysregulated in *N1*<sup>+/-</sup> ECs, whereas in shear stress conditions, 791 mRNAs were altered (with approximately half being dysregulated in both conditions) (Figures 3D and 3G and Table S5). GO analysis showed that NOTCH and DELTA-NOTCH signaling were among the top pathways dysregulated in *N1* haploinsufficiency (shear and/or static conditions), indicating that the *N1*<sup>+/-</sup> iPSC-derived ECs were indeed modeling a defect in *N1* activity (Figure 3E and Table S6). Furthermore, top dysregulated GO pathways included endochondral ossification and inflammatory response, two pathways thought to play a major role in CAVD. In addition, 180 small non-coding RNAs (ncRNAs) were significantly dysregulated in *N1*<sup>+/-</sup> ECs, and these were enriched for miRNAs and CDBox RNAs (Figures S5B and S5C). Surprisingly, no ncRNAs were significantly dysregulated under shear stress, indicating that shear-stress-induced *N1* signaling may be sufficient to restore ncRNA networks in *N1*<sup>+/-</sup> ECs to their WT state, similar to the shear-stress-induced restoration of a subset of dysregulated mRNAs. ncRNAs altered in *N1*<sup>+/-</sup> ECs reflected similar processes as observed for dysregulated mRNAs. Anti-osteogenic (e.g., miR-20a, 26a, 30e, and 106a) and anti-atherogenic miRNAs (e.g., miR-126) were downregulated in *N1*<sup>+/-</sup> ECs, whereas pro-osteogenic (e.g., miR-30d) and pro-atherogenic (e.g., miR-663) miRNAs were upregulated (Goettsch et al., 2013; Li et al., 2013; Schober et al., 2014).

When we tested whether unique features of the shear stress condition were dysregulated in *N1* heterozygosity, we found that *N1*<sup>+/-</sup> ECs did not properly activate anti-calcific genes normally induced by shear stress (Figures 3D and 3F). In total, 30% of shear-responsive genes in WT ECs were dysregulated



**Figure 3. Gene Networks Dysregulated in *N1* Haploinsufficient Isogenic iPSC-Derived ECs**

(A) Pedigrees of two families affected with congenital heart disease and valve calcification due to *N1* mutations. Squares, males; circles, females.

(B) mRNA expression of *N1* and compensatory upregulation of *NOTCH4*.

(C) mRNA expression of canonical *N1* targets *HES1* and *EFNB2*.

(D) Log<sub>2</sub> fold change in mRNA expression in *N1*<sup>+/-</sup> versus WT ECs in static and shear stress conditions of 1,303 genes significantly dysregulated in *N1*<sup>+/-</sup> ECs.

(E) Top GO pathways enriched among genes dysregulated in *N1*<sup>+/-</sup> ECs.

(F) Examples of anti-osteogenic (*GREM1* and *DKK1*), antioxidant (*TXNRD1*), and anti-atherogenic (*CYP1B1*) shear-responsive genes not properly activated in *N1*<sup>+/-</sup> ECs.

(G) Overlap of statistically significant gene sets.

In (B–G): WT n = 3, *N1*<sup>+/-</sup> n = 2 (isogenic ECs); error bars represent SE; \*p < 0.05 by negative binomial test with FDR correction. See also Figures S2, S3, S4, and S5 and Tables S5 and S6.



in  $N1^{+/-}$  ECs and were thus N1 dependent (Figure 3G).  $N1^{+/-}$  ECs showed downregulation of shear-specific antagonists of pro-osteogenic BMP and WNT pathways, including genes producing secreted proteins GREM1 and DKK, respectively (Van Handel et al., 2012). Furthermore, shear-exposed  $N1^{+/-}$  ECs failed to upregulate anti-atherogenic factors such as *CYP1B1* and showed aberrant upregulation of pro-inflammatory genes, including *IRF6* (Conway et al., 2009; Kwa et al., 2014). In both static and shear stress conditions,  $N1^{+/-}$  ECs showed an increase in cell-cycle genes such as *CDC20* and *CDCA2* (with additional genes such as *CCNA1* and *CCND1* activated in static conditions). Furthermore, we observed downregulation of *PDE3A* and upregulation of *PDE2A*, which would be predicted to alter EC permeability and promote inflammatory cell infiltration (Surapisitchat et al., 2007). Finally, when exposed to shear stress,  $N1^{+/-}$  ECs did not appropriately upregulate genes involved in the oxidative stress response such as *CYGB* and *TXNRD1* (Li et al., 2007) and showed downregulation of MMPs such as *MMP7* and *9* (with additional MMPs *10*, *19*, and *24* downregulated in static conditions) (Figure S5E). Overall, N1 haploinsufficient ECs could not mediate the normal anti-calcific response induced by shear stress and showed aberrant upregulation of pro-osteogenic and inflammatory signaling.

### Epigenetic Dysregulation Correlated with Pro-calcific Gene Expression in $N1^{+/-}$ ECs

To determine whether decreased dosage of N1 perturbed the epigenetic state of genes dysregulated in  $N1^{+/-}$  ECs, we performed genome-wide ChIP-seq for H3K4me3, H3K27ac, H3K4me1, and H3K27me3 in WT or  $N1^{+/-}$  patient-specific iPSC-derived ECs under static and shear stress conditions. We determined the density of the proximal promoter mark H3K4me3 within 3 kb of TSSs of genes dysregulated in  $N1^{+/-}$  ECs. Given that H3K27ac, H3K4me1, and H3K27me3 mark both proximal and distal regulatory domains, we determined the density of these marks within 15 kb of TSSs of dysregulated genes.

When we clustered dysregulated genes based on their expression and epigenetic state in  $N1^{+/-}$  ECs compared to WT ECs, the resulting clusters defined functional groups of genes involved in distinct aspects of CAVD (Figure 4A). For example, cluster I contained genes involved in ECM and pro-osteogenic WNT signaling, which showed increased transcription in both static and shear stress conditions without much change in any of the four evaluated chromatin marks. However, cluster J distinguished genes involved in pro-osteogenic BMP signaling whose transcriptional upregulation was accompanied by increased H3K4me3 and H3K27ac activating marks in both static and shear stress conditions, as well as decreased H3K27me3 repressive marks in the static condition. Thus, functional groups of genes shared common epigenetic mechanisms associated with transcriptional dysregulation in  $N1^{+/-}$  ECs.

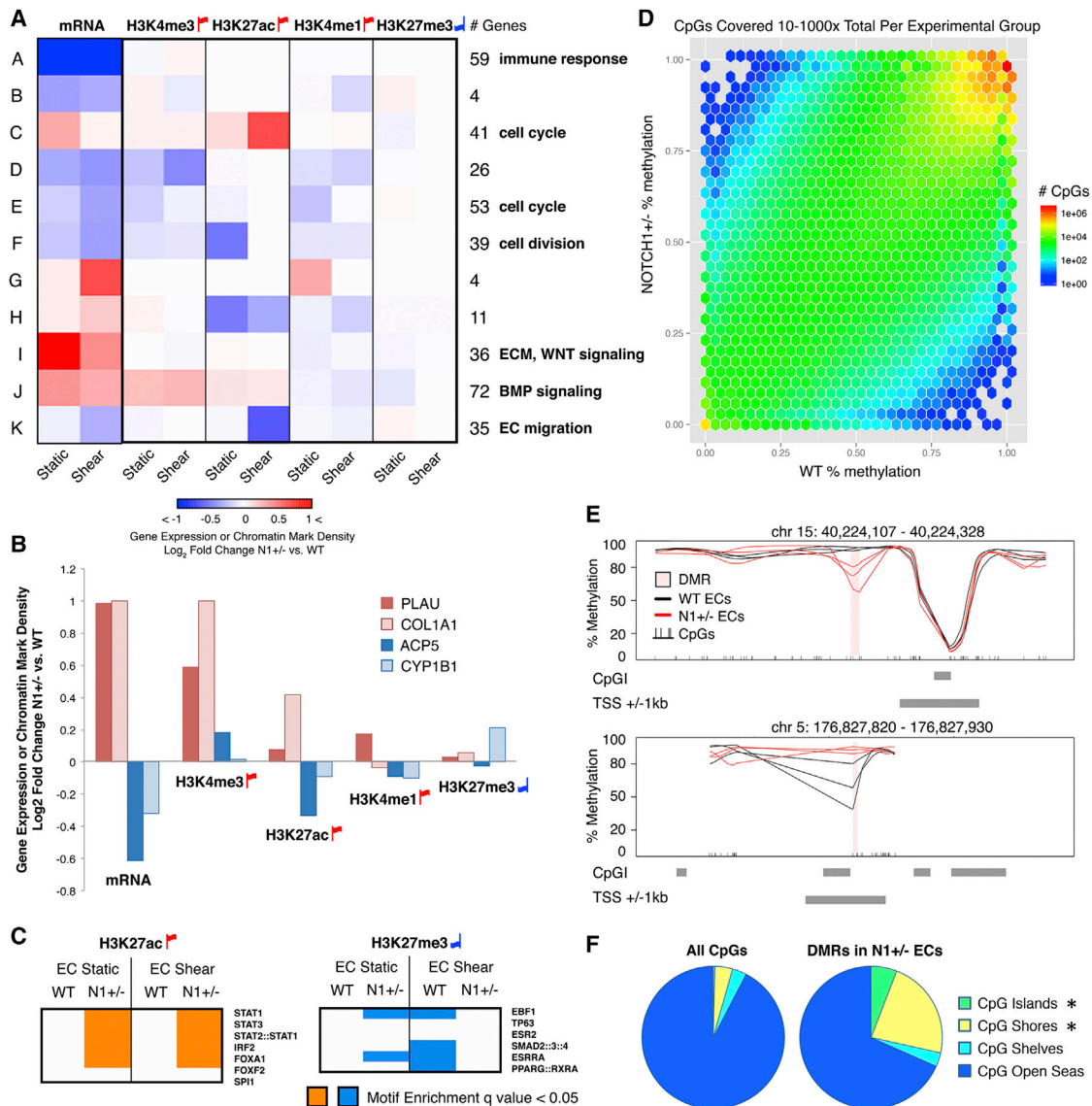
Focusing on individual genes revealed that histone modification changes correlated with pro-calcific gene dysregulation in  $N1^{+/-}$  ECs (Figure 4B). Pro-osteogenic genes involved in endochondral ossification (e.g., *PLAU*) and osteoblast regulation (e.g., *COL1A1*) showed upregulated mRNA expression associated with increased activating marks H3K4me3, H3K27ac, and

H3K4me1 (Engelholm et al., 2001). Conversely, transcriptionally downregulated anti-calcific factors, including anti-atherogenic genes such as *CYP1B1* and osteoclast genes such as *ACP5*, showed decreased H3K27ac activating marks, as well as increased repressive H3K27me3 marks (Alatalo et al., 2000).

We next investigated whether epigenetic dysregulation in  $N1^{+/-}$  ECs reflected changes in the distribution of key TF motifs within active and repressed regions (Figure 4C). Epigenetic studies in WT ECs above showed that pro-inflammatory STAT and IRF motifs were enriched in H3K27ac-marked enhancers only in static conditions. Notably, STAT and IRF motifs were significantly enriched in H3K27ac-marked activation sites in  $N1^{+/-}$  as compared to WT ECs during both static and shear stress conditions (Figure 4C). This suggests that in  $N1^{+/-}$  ECs there is an even greater increase in STAT and IRF signaling in static conditions and that these pro-inflammatory signals are not appropriately downregulated in response to shear stress. Conversely, H3K27me3-marked repressive domains in WT ECs were enriched for SMAD2/3/4 motifs in shear stress conditions, whereas H3K27me3 sites in  $N1^{+/-}$  ECs were depleted for SMAD2/3/4 motifs in shear stress, suggesting a de-repression of pro-osteogenic TGF $\beta$  signaling as a consequence of N1 haploinsufficiency. Furthermore, in both static and shear stress conditions, H3K27ac sites in  $N1^{+/-}$  ECs were enriched for RUNX1 motifs compared to WT ECs, likely indicating a progression to early chondrogenic signaling given RUNX1's role in chondroblasts and bone formation (Smith et al., 2005; Yamashiro et al., 2004) in addition to its role in hematopoiesis (Okuda et al., 1996). In sum,  $N1^{+/-}$  ECs revealed a shift of activated chromatin marks toward putative pro-inflammatory and osteogenic regulatory domains and a depletion of repressive marks from putative pro-osteogenic enhancers.

To understand whether  $N1^{+/-}$  ECs underwent more permanent silencing of regulatory regions that may in principle be related to disease, we evaluated the DNA methylation landscape in WT and  $N1^{+/-}$  ECs by whole-genome bisulfite sequencing. We concentrated on the static condition as preliminary studies showed negligible change in methylation status in response to shear stress assays, which occurred over only 24 hr with little cell division to allow for methylation turnover. Bisulfite sequencing revealed 248 differentially methylated regions (DMRs) in  $N1^{+/-}$  ECs compared to WT ECs (Figures 4D and 4E). Nearly half of these DMRs were novel regions not identified in previous efforts to capture the dynamic DNA methylation landscape essential for normal development (Ziller et al., 2013). This suggests that dysregulation of DNA methylation in disease settings may partially occur in regulatory domains otherwise stable throughout development.

DMRs due to N1 haploinsufficiency were significantly enriched for CpG islands (CpGIs) and shores ( $\pm 2$  kb from CpGIs) and depleted for CpG open seas ( $>4$  kb from CpGIs) (Figure 4F). The largest enrichment of DNA methylation changes occurred at CpG shore regions, suggesting that these regions might be the least stable in the disease state. In many regions, changes in DNA methylation were accompanied with chromatin mark dysregulation. Regions hypermethylated in  $N1^{+/-}$  ECs lost H3K4me3 or H3K27ac activating marks present in WT ECs, whereas regions hypomethylated in  $N1^{+/-}$  ECs gained



**Figure 4. Epigenetic Dysregulation in  $N1^{+/−}$  ECs**

(A) Hierarchical clustering of genes based on  $\log_2$  fold change of expression and enrichment of histone modifications within 3 kb (H3K4me3) or 15 kb (H3K27ac, H3K4me1, and H3K27me3) of the TSS in  $N1^{+/−}$  versus WT ECs. Enriched GO pathways within each cluster are shown on the right.

(B) Mean  $\log_2$  fold change in  $N1^{+/−}$  versus WT static ECs of mRNA expression and histone modifications as in (A) of individual pro-osteogenic (PLAU and COL1A1), osteoclast (ACP5), and anti-atherogenic (CYP1B1) genes.

(C) TF motif enrichment in  $N1^{+/−}$  versus WT chromatin marks in static or shear stress conditions. Motifs tested were drawn from unique clusters identified in Figure 1D.

(D) Relative mean DNA methylation of CpGs in  $N1^{+/−}$  (vertical axis) versus WT (horizontal axis) ECs in static conditions. Plot includes only CpGs with 10–1,000× total sequence coverage between three biological replicates per experimental group.

(E) Examples of the 248 DMRs identified in  $N1^{+/−}$  versus WT ECs.

(F) Distribution of DMRs or all CpGs relative to CpGIs. Shores are <2 kb flanking CpGIs; shelves are <2kb flanking outward from shores; open seas are >4 kb flanking CpGIs. \* $p < 0.05$  by  $\chi^2$  test with Bonferroni correction.

In (A–C): WT  $n = 5$ ,  $N1^{+/−}$   $n = 3$  (patient-specific ECs). Red up flags, activating marks; blue down flags, repressive marks. In (D–F): WT  $n = 3$ ,  $N1^{+/−}$   $n = 3$  (patient-specific ECs). See also Figure S6 and Table S7.

H3K4me3 or H3K27ac marks (Figure S6A). The extensive conversion of these DMRs between the silenced and activated state suggests robust changes in the epigenetic landscape of  $N1^{+/−}$  ECs.

### N1 Binding Sites Showed Dysregulation in H3K27ac Chromatin Marks in $N1^{+/−}$ ECs

To discern the aspects of transcriptional and epigenetic dysregulation in  $N1^{+/−}$  ECs directly associated with N1 genome

occupancy, we performed endogenous N1 ChIP-seq in primary human aortic ECs (HAECs). We found that 414 of the 1,303 genes dysregulated in  $N1^{+/-}$  ECs showed N1 binding in proximity to their TSSs (within 20 kb) as determined by peaks called in the N1 ChIP-seq (Figure 5A). Overall, dysregulated genes were significantly more likely to be found in proximity to a N1 peak than non-dysregulated genes ( $p < 0.05$ ). Genes with the most significant N1 ChIP peaks near their TSSs were most often downregulated in  $N1^{+/-}$  ECs, which is consistent with N1 acting as a transcriptional activator. However, many putative direct targets were upregulated in  $N1$  heterozygosity, indicating that N1 might also directly repress pro-calcific genes in WT ECs.

Most N1 binding events occurred in distal intergenic regions, but binding was most significantly enriched in promoter regions within 1 kb from TSSs compared to genomic background (Figure 5B). Putative direct N1 targets whose expression was unchanged by N1 haploinsufficiency on average showed greater enrichment of N1 binding near the TSS than targets dysregulated in  $N1^{+/-}$  ECs (Figure 5C). Sites bound with lower levels of N1 may be more sensitive to a reduction in N1 dose, as they may be closer to a critical occupancy threshold required for gene activation. To determine motifs enriched in N1 binding sites, we compared motifs in genomic regions  $\pm 25$  base pairs (bps) from N1 peak summits to motifs found in general open chromatin regions within ECs as determined by H3K27ac ChIP-seq (Figure 5D). Only 17% of N1 peaks contained the motif bound by its canonical DNA-binding partner CSL, suggesting that N1 may target DNA through alternate binding partners in ECs. However, N1-bound sites were enriched for motifs, including RUNX1 motifs, as previously seen in T-lymphoblastic leukemia cells (Wang et al., 2011), as well as IRF and STAT motifs. Because RUNX1, IRF, and STAT motifs were all enriched in enhancers active specifically in  $N1^{+/-}$  ECs, N1 complex binding may sufficiently compete with the binding of these TFs in the WT, but not mutant, state to prevent activation of related pro-osteogenic and pro-inflammatory signaling pathways.

We next investigated whether histone modifications near direct N1 binding sites were altered in ECs with decreased dosage of N1. In WT ECs, genomic regions  $\pm 1$  kb from N1 peak summits were significantly enriched for H3K27ac and H3K4me1 and depleted, though not significantly, for H3K27me3 as defined by ChIP-seq (Figure 5E). In contrast, genomic regions  $\pm 1$  kb from N1 peak summits showed significant alterations in H3K27ac and to a lesser degree in H3K4me1 in  $N1^{+/-}$  ECs in both static and shear stress conditions compared to WT ECs (Figures 5F and S6B).  $N1^{+/-}$  ECs showed increased H3K27ac at some sites and decreased H3K27ac at other sites, suggesting that N1 binding has a locus-specific effect on H3K27ac. For example, decreased H3K27ac surrounding two N1 binding sites within the gene body of the previously undescribed target *ARHGEF17* correlated with its significant transcriptional downregulation in static conditions (Figure 5G).

We clustered N1 binding sites by the change in H3K27ac in N1 haploinsufficiency compared to WT, which revealed that the effect of shear stress on H3K27ac at N1 binding sites was reduced in  $N1^{+/-}$  ECs (Figures 5F and 5H). Within N1-bound sites where shear stress normally increased acetylation (Cluster B),

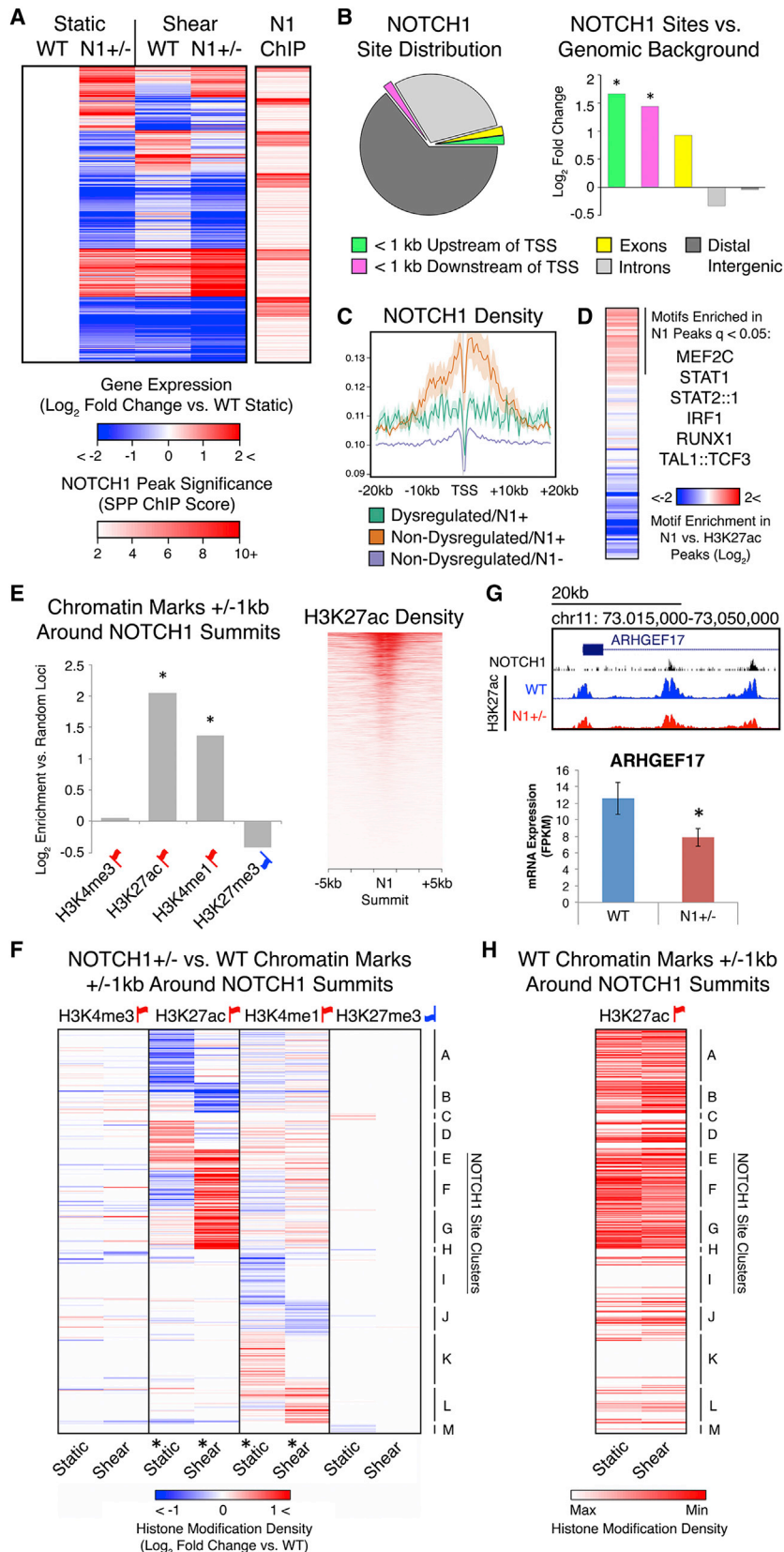
mutant ECs showed less acetylation in shear stress than WT ECs. Conversely, sites with decreased acetylation in shear stress (Clusters E–G) showed incomplete deacetylation in mutant ECs as compared to WT ECs. This dampening of shear stress effects was less pronounced in distal H3K27ac loci devoid of N1 binding sites (Figure S6C). However, the pattern was present in H3K27ac loci proximal (within 1 kb) to TSSs and distinguished functional groups of genes (Figure S6D and Table S7). For example, promoters of genes encoding inhibitors of pro-osteogenic WNT signaling (e.g., *HBP1* and *TLE*) (Sampson et al., 2001; Wu et al., 2014) normally experienced increased H3K27ac in response to shear stress but showed lower levels of acetylation in  $N1^{+/-}$  ECs as compared to WT ECs in shear stress conditions. Because almost all N1 binding sites showing H3K27ac dysregulation were distal from gene TSSs, these data suggest that N1 binding may mediate the effect of shear stress on H3K27ac specifically at distal enhancers.

### Manipulating Dysregulated Regulatory Nodes Restores Expression of N1 Downstream Targets toward WT Levels

We sought to harness our knowledge of the transcriptional and epigenetic dysregulation caused by N1 haploinsufficiency to identify putative regulatory nodes within the gene network downstream of N1 that might serve as therapeutic targets. We employed a network inference algorithm to predict network connections using RNA-seq data of WT and  $N1^{+/-}$  ECs in static and shear stress conditions (Margolin et al., 2006). With N1 as a network hub, we predicted several genes to be directly connected to N1 (Figure 6A). These included the canonical targets *EPHNB2* and *HES4*, as well as *ARHGEF17*, which we described above as a direct N1 target dysregulated in  $N1^{+/-}$  ECs. Genes with putative direct connections to N1 were themselves highly interconnected, suggesting that these genes might co-regulate one another to provide additional network stability. Each of these genes further branched out to connect to particular sets of genes, such as *HES4*, which connected to many genes strongly dysregulated in  $N1^{+/-}$  ECs, and *ARHGEF17*, which connected to genes that were nearly all highly shear-responsive (Figures S7A and S7B).

When we expanded our network prediction to include all dysregulated genes as hubs, we obtained a gene network with scale-free properties illustrating the predicted connections between genes downstream of N1 (Figure 6B). Although most genes were connected to few dysregulated targets, several genes encoding transcriptional regulators were connected to a large portion of the genes dysregulated in N1 haploinsufficiency, suggesting that these putative regulatory nodes may control the majority of altered genes (Figures 6B and 6C). Among the nodes connected to the most dysregulated genes were *SOX7*, the WNT signaling effector *TCF4*, and the BMP signaling effector *SMAD1*, all upregulated in  $N1^{+/-}$  ECs.

Using siRNAs, we corrected the aberrant upregulation of *SOX7*, *TCF4*, or *SMAD1*, alone or in combination, in an effort to restore the normal EC gene network. We monitored the effects of these perturbations on the initial putative targets (*RASSF4*, *THSD1*, *ACE*, *PDE2A*, and *GREM1*) selected based on connections predicted in our inferred gene network (Figure 6D). When



**Figure 5. Transcriptional and Epigenetic Dysregulation Directly Associated with N1 Genome Occupancy**

(A) Left: K means clustering of putative direct N1 targets defined as genes significantly dysregulated in *N1*<sup>+/-</sup> ECs with N1 ChIP peaks within 20 kb of the TSS. Right: significance of N1 peaks within 20 kb of the TSS of 414 putative direct N1 targets.

(B) Left: distribution of N1 peaks. Right: log<sub>2</sub> fold change of proportion of N1 peaks versus genomic background in indicated regions. \*p < 0.05 by X<sup>2</sup> test with Bonferroni correction.

(C) N1 density around the TSS of genes dysregulated in N1 haploinsufficiency with N1 peaks within 20 kb (green), non-dysregulated genes with N1 peaks within 20 kb (orange), or non-dysregulated genes without N1 peaks within 20 kb (blue).

(D) Motifs significantly enriched (q < 0.05 by motifDiverge with FDR correction) within 25 bps of N1 peak summits compared to H3K27ac peaks in ECs in static conditions.

(E) Left: Log<sub>2</sub> fold change of overlap of chromatin marks in WT ECs with 1 kb around N1 summits versus random non-gap genomic loci. \*p < 0.05 by X<sup>2</sup> test with Bonferroni correction. Right: H3K27ac density near N1 summits.

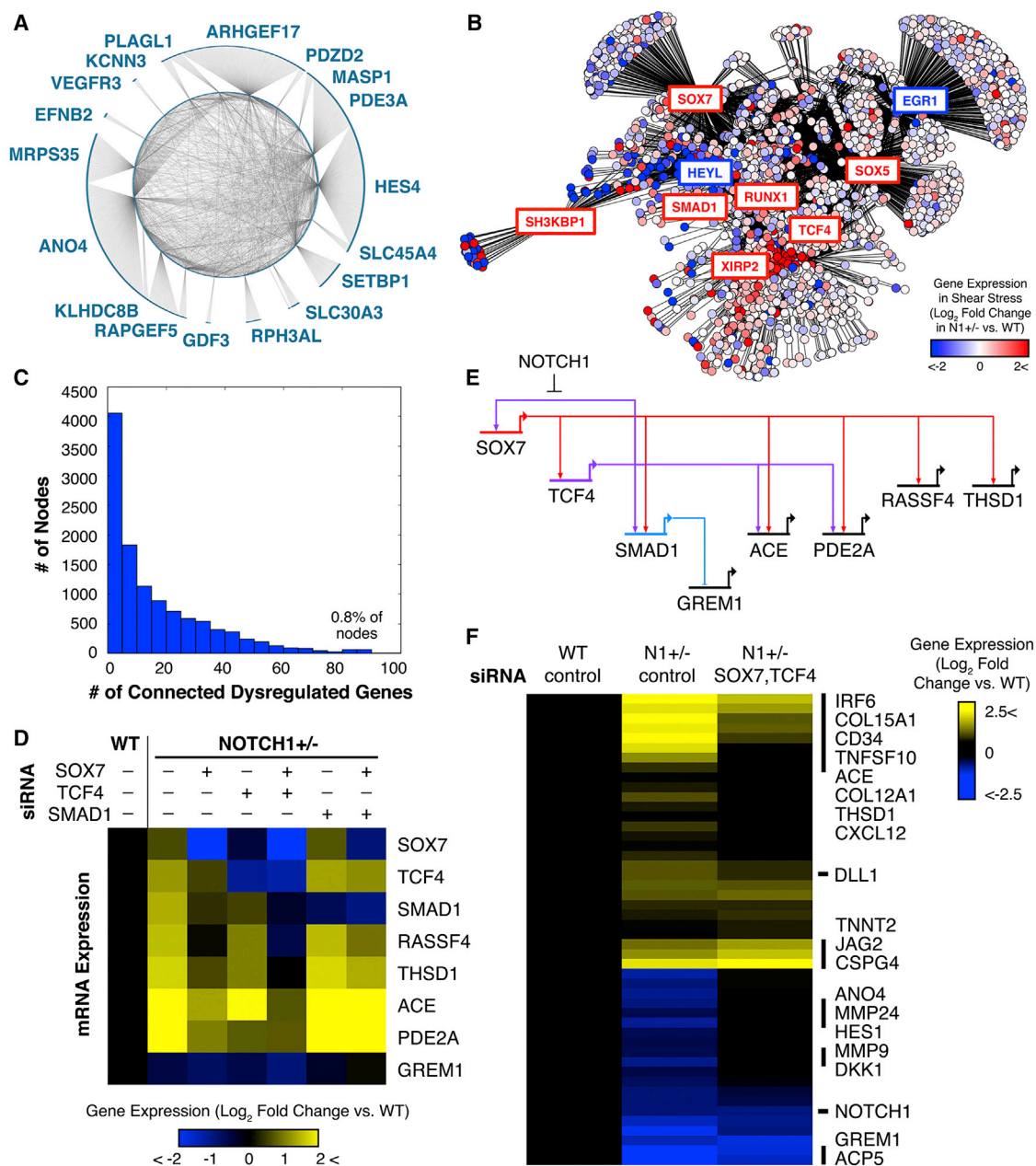
(F) Hierarchical clustering based on log<sub>2</sub> fold change of *N1*<sup>+/-</sup> versus WT histone modification density within 1 kb of N1 summits. \*p < 0.05 by KS test with Bonferroni correction (histone modification dysregulation around N1 summits versus random non-gap genomic loci).

(G) Top: N1 peaks and WT or *N1*<sup>+/-</sup> H3K27ac near *ARHGEF17*. Bottom: mean mRNA expression of *ARHGEF17*. Error bars represent SE; \*p < 0.05 by negative binomial test with FDR correction.

(H) Relative H3K27ac density within 1 kb of N1 summits ordered as in (F) in WT ECs in static or shear stress conditions.

In (A–H): gene expression: WT n = 3, *N1*<sup>+/-</sup> n = 2 (isogenic iPSC-derived ECs). Chromatin marks: WT n = 5, *N1*<sup>+/-</sup> n = 3 (patient-specific iPSC-derived ECs). N1 genome occupancy: WT n = 1 (union of three technical replicates) (primary HAECs). Red up flags, activating marks; blue down flags, repressive marks. See also Figure S6.





**Figure 6. Manipulation of Dysregulated Regulatory Nodes to Restore the EC Gene Network**

(A) Putative regulatory nodes directly connected to N1 in the predicted network and their interconnections ( $p < 0.05$ ).

(B) Predicted gene regulatory network in ECs ( $p < 0.05$ ) with each circle representing a gene and color indicating  $\log_2$  fold change of  $N1^{+/-}$  versus WT expression in shear stress conditions. Boxed genes are putative dysregulated regulatory nodes with red and blue boxes indicating up- or downregulated genes, respectively.

(C) Histogram of number of nodes with different numbers of connected dysregulated genes. A small number of master regulators may control the majority of dysregulated genes.

(D) Effect of control, SOX7, TCF4, and/or SMAD1 siRNA on  $N1^{+/-}$  EC mRNA expression of indicated genes as detected by QPCR.

(E) Gene regulatory subcircuit assembled based on perturbation results and network prediction.

(F) Effect of combined SOX7 and TCF4 siRNA on restoring  $N1^{+/-}$  versus WT expression of 48 genes dysregulated in N1 haploinsufficiency.  $N = 2$ . See also Figure S7.

we intervened solely with SOX7 siRNA, both TCF4 and SMAD1 were restored toward their WT expression levels, indicating that SOX7 is upstream of these predicted nodes. In addition, the expression of all other putative targets except GREM1 was partially to fully restored to WT levels. Intervening with TCF4 siRNA also reduced the aberrant upregulation of SOX7,

demonstrating a positive feedback loop between *SOX7* and *TCF4*. *SMAD1* also shifted toward its WT expression, either through direct regulation by *TCF4* or indirect regulation through *SOX7*. Knockdown of *TCF4* also partially restored the expression of most putative downstream targets tested (*RASSF4*, *THSD1*, and *PDE2A*), albeit to a lesser degree than knockdown of *SOX7*. However, treating *N1*<sup>+/-</sup> ECs with a combination of *SOX7* and *TCF4* siRNA had even more dramatic effects in restoring the WT expression of all putative targets, excluding *GREM1*. In contrast, knocking down *SMAD1* restored *GREM1* expression toward WT levels without affecting *SOX7*, *TCF4*, or any of the other putative targets. Ultimately, based on our perturbation experiments and network inference, we were able to identify a gene network sub-circuit involving nodes *SOX7*, *TCF4*, and *SMAD1* that controls downstream gene dysregulation in *N1*<sup>+/-</sup> ECs (Figure 6E). *N1* is responsible for repressing this sub-circuit from activation in WT ECs while decreased *N1* levels in *N1*<sup>+/-</sup> ECs are insufficient to prevent its activation.

To determine whether intervening with the *SOX7* and *TCF4* regulatory nodes had more widespread effects on restoring the gene network downstream of *N1*, we expanded our panel to include 48 genes dysregulated in *N1*<sup>+/-</sup> ECs selected based on our inferred gene network and RNA-seq data (Figure 6F). We found that the combination of *SOX7* and *TCF4* siRNA restored the expression of the majority of these genes toward the WT state. *N1* expression was unaffected, but expression of its canonical downstream target *HES1* was restored toward WT, suggesting possible repair of the pathways downstream of *N1* signaling. The siRNA treatment alleviated the downregulation of genes encoding MMPs (MMP24, MMP9), which may serve to degrade ECM in the valve, and secreted anti-osteogenic factors such as the WNT inhibitor DKK1 that may help prevent calcification in underlying VICs. Additionally, the treatment reduced upregulation of pro-inflammatory genes such as *IRF6* and cytokine *CXCL12*, as well as collagens *COL15A1* and *COL12A1* that may contribute to calcification. Overall, correcting the aberrant upregulation of regulatory nodes *SOX7* and *TCF4* using siRNA was able to restore expression of genes within the network dysregulated in *N1*<sup>+/-</sup> ECs toward the WT state.

## DISCUSSION

We have defined the mechanisms critical for normal human EC differentiation and response to shear stress, determined how these mechanisms are perturbed in *N1* haploinsufficient cells, and intervened at key regulatory nodes to restore *N1*<sup>+/-</sup> ECs toward their WT state. Overall, iPSC-based modeling of human *N1* mutations allowed a rigorous interrogation of the gene networks disrupted in ECs from patients with CAVD to reveal novel targets for intervention. Moreover, the findings here demonstrate mechanisms by which dose-reduction of a TF can alter the epigenetics and transcriptome in a human disease model.

### Transcriptional and Epigenetic Mechanisms Governing EC Differentiation and Response to Shear Stress

Among the more interesting findings from transcriptional and epigenetic profiling during EC differentiation was that shear

stress induced a highly coordinated suppression of pro-osteogenic and inflammatory signaling in ECs that may be critical to protect against calcification events. Shear-dependent histone modifications correlated with upregulation of anti-osteogenic genes, including those encoding secreted BMP and WNT antagonists that may represent a paracrine signaling method for shear-exposed ECs to prevent the calcification of neighboring tissue. Synchronously, shear stress downregulated pro-inflammatory cytokines, reduced STAT and IRF signaling effectors characteristic of ECs in static conditions, and shifted activating histone modifications toward anti-inflammatory and antioxidant motifs. It is interesting to consider that ECs throughout the vasculature may function to repress pro-osteogenic events in response to laminar shear stress, given the propensity for calcification at sites of vascular bifurcation experiencing turbulent blood flow.

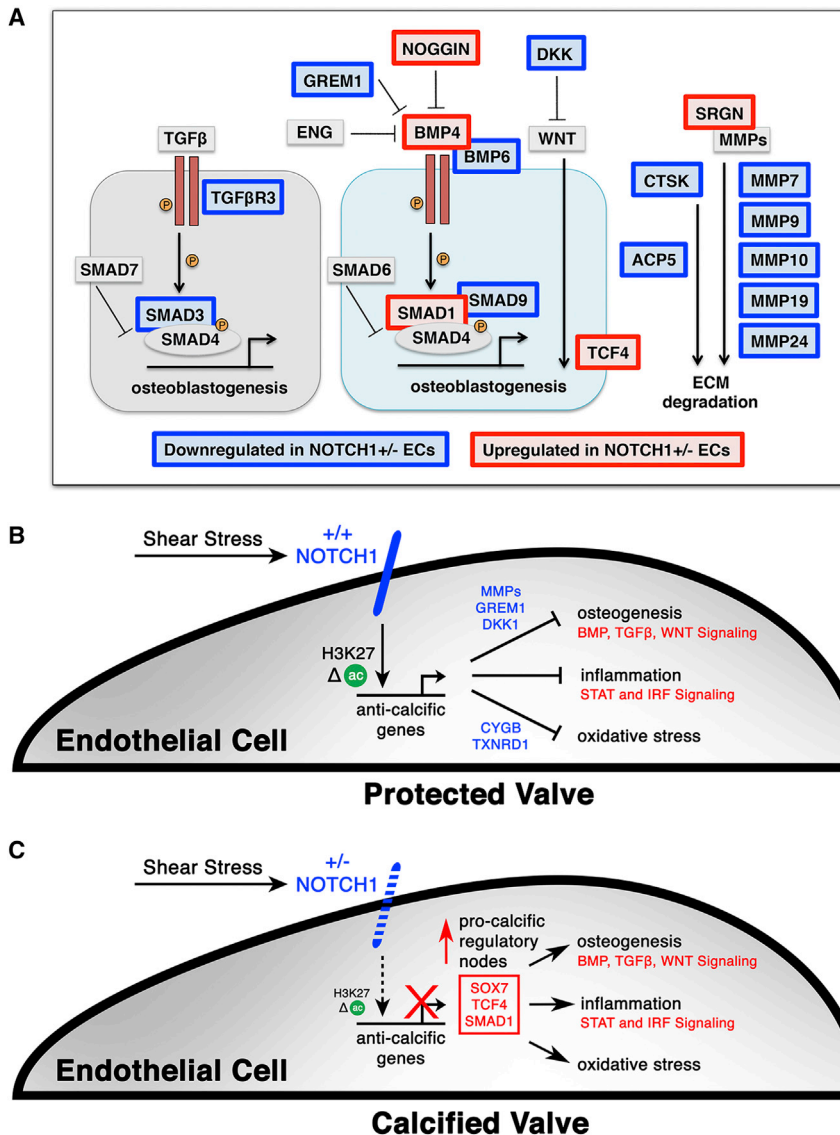
### Transcriptional Consequences of *N1* Haploinsufficiency in iPSC-Based Modeling of CAVD

iPSC-based modeling of human *N1* mutations in CAVD revealed that *N1* haploinsufficiency disrupts the appropriate EC response to shear stress (Figures 7A–7C). In contrast to WT ECs, shear-exposed *N1*<sup>+/-</sup> ECs failed to upregulate anti-osteogenic factors, including secreted BMP and WNT antagonists that may be critical for preventing calcification of underlying VICs. They instead overexpressed pro-osteogenic genes such as *BMP4*, suggestive of an osteoblast-like switch, highlighting the importance of the *N1*-dependent response to shear stress in maintaining the cell fate of valve ECs. Conversion of valve ECs into osteoblast-like cells has been reported in disease states (Hofmann et al., 2012), which is consistent with *N1* functioning to repress this aberrant gene program. Interestingly, the secreted anti-osteogenic factor, matrix Gla protein (MGP), was very lowly expressed in iPSC-derived ECs despite its abundance in native murine valve tissue (Luo et al., 1997), making it difficult to determine whether MGP was a shear-responsive *N1* target in human cells. Nevertheless, the transcriptional disturbances in *N1*<sup>+/-</sup> ECs indicated a dysregulated inflammatory environment and vulnerability to oxidative stress that may fuel the progression of calcification in an aortic valve without the anti-calcific barriers normally erected by shear-exposed ECs.

Determining the pathways dysregulated in *N1*<sup>+/-</sup> ECs focuses therapeutic efforts on reinstating the proper EC response to shear stress and restoring the anti-osteogenic and anti-inflammatory barriers against calcification in the valve. Furthermore, defining the factors important for maintaining these barriers provides insight into alternative genes that may be mutated in CAVD patients without mutations in *N1*. Future work delineating whether mutations in various members of the network preventing CAVD lead to aortic valve disease may explain why only a portion of patients with BAV progress to valve calcification.

### *N1* Genome Occupancy and Epigenetic Dysregulation in *N1*<sup>+/-</sup> ECs

Determining the transcriptional and epigenetic consequences of *N1* haploinsufficiency occurring directly at *N1*-bound sites



**Figure 7. Model of Mechanisms Regulating Pro-calcific Events in N1 Haploinsufficient ECs**

(A) Diagram of osteogenic pathways dysregulated in N1 haploinsufficiency. Red indicates upregulation in  $N1^{+/-}$  ECs and blue indicates downregulation in  $N1^{+/-}$  ECs.

(B) Model of WT ECs. Shear stress activates N1 signaling in ECs, leading to epigenetic changes at N1-bound enhancers and transcriptional activation of anti-calcific gene programs that prevent osteogenesis, inflammation, and oxidative stress to protect the valve from calcification.

(C) Model of  $N1^{+/-}$  ECs, which cannot mediate the proper response to shear stress, leading to epigenetic dysregulation at N1-responsive enhancers and aberrant upregulation of pro-calcific regulatory nodes.

state and transcriptional regulation of the most sensitive direct targets.

N1 haploinsufficiency had broad downstream effects on the epigenetic landscape in ECs. Epigenetic dysregulation extended to DNA methylation, indicating an extensive shift in the regulatory state of key domains in ECs. CpG shore regions were the most vulnerable to dysregulation, showing the largest enrichment of methylation changes in N1 haploinsufficient ECs. These data support the notion that CpG shores are the least stable in the disease state, which is consistent with previous reports of shores displaying the most methylation differences in the context of cancerous cells and specific tissue types (Irizarry et al., 2009). This ability to differentiate between distinct cell states may have important implications in diagnosis of disease. For example, DNA methylation

provides insight into how TF dosage differentially affects targets leading to human disease. At baseline in WT ECs, gene targets dysregulated in  $N1^{+/-}$  ECs had lower N1 occupancy proximal to their TSS compared to non-dysregulated targets. Thus, genes most sensitive to decreased dosage of N1 begin with lower levels of N1 binding at baseline, potentially placing them closer to the threshold at which binding becomes insufficient to affect transcription.

Consistent with previous reports of N1 recruitment of histone acetyltransferases (Yashiro-Ohtani et al., 2014), N1 binding sites showed the greatest changes in H3K27ac, compared to other epigenetic marks, and could not mount the proper epigenetic response to shear stress in  $N1^{+/-}$  ECs. The changes in H3K27ac correlated with transcriptional dysregulation of putative direct targets such as *ARHGEF17*, which we found to be a major predicted regulatory node, indicating that WT levels of N1 binding are required to maintain the appropriate epigenetic

may provide a useful marker for patients who are at risk for valve calcification.

#### Identification of Dysregulated Transcriptional Nodes

An incomplete understanding of the molecular mechanisms involved in the development of CAVD has hampered the design of effective therapies. By thoroughly interrogating the transcriptional and epigenetic consequences of N1 haploinsufficiency, we identified transcriptional nodes controlled by SOX7 and TCF4, where intervention impacted a large portion of the gene dysregulation in patient-specific ECs. The ability to exert a concerted influence on gene networks disrupted in cells from patients with CAVD by targeting discrete regulatory nodes has potential therapeutic implications. Using this iPSC-based disease model, we anticipate screening for small molecules that target central regulatory nodes to inhibit or delay the progression of calcification in patients at risk for CAVD.

The work presented here demonstrates the value of computationally integrating genome-wide transcriptome, DNA methylation, and histone modification data with gene network analyses to reveal the consequences of human disease-causing mutations. We believe this type of broad and comprehensive approach will serve as the foundation for rational drug design for many disorders in the coming years.

## EXPERIMENTAL PROCEDURES

Experimental details can be found in [Extended Experimental Procedures](#).

## ACCESSION NUMBERS

The GNomEx experiment numbers for the data reported in this paper are as follows: mRNA-seq on patient-specific lines: 196R; mRNA-seq on isogenic lines: 351R; ncRNA microarrays: 194R; histone modification; ChIP-seq: 195R2; bisulfite sequencing: 320R; and N1 ChIP-seq: 355R1.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.035>.

## AUTHOR CONTRIBUTIONS

C.V.T. designed/performed experiments and designed/performed bioinformatic analyses on transcriptional, epigenetic, genome occupancy, and network inference studies. M.L. contributed to bioinformatic analyses on epigenetic data. M.P.W. supervised iPSC reprogramming and provided EC differentiation method. C.V.T. and L.L. performed EC differentiations and TALEN genome engineering. C.V.T. and D.H. performed ChIP assays. K.S.P. supervised computational design, analyses, and interpretation and edited the manuscript. B.G.B. supervised epigenetic analyses and interpretation and edited the manuscript. D.S. designed experiments and analyses and supervised the work. C.V.T. and D.S. wrote the manuscript.

## ACKNOWLEDGMENTS

We are grateful to members of the Srivastava Lab and Gladstone/UCSF community for helpful scientific discussions. We thank the Gladstone Genomics, Histology, and Stem Cell Cores and UCSF ESC Targeting Core for their technical expertise. We thank G. Howard and B. Taylor for their editorial help and K. Ivey and C. Gifford for critical comments. This work was supported by NIH/NHLBI grants U01 HL098179 (B.G.B., K.S.P., and D.S.) and U01 HL100406 (B.G.B. and D.S.). D.S. was supported by the L.K. Whittier, William H. Younger Family, and Eugene Roddenberry Foundations and the California Institute for Regenerative Medicine. B.G.B. was supported by the Lawrence J. and Florence A. DeGeorge Charitable Trust/American Heart Association (AHA) Established Investigator Award, William H. Younger Family Foundation, and a grant from the California Institute for Regenerative Medicine (RB4-05901). K.S.P. was supported by a grant from NIH/NHLBI (HL089707). C.V.T. was supported by the AHA Predoctoral, UCSF Discovery, Roddenberry, and Winslow Fellowships and the UCSF Core Exploratory Award, Developmental and Stem Cell Biology Graduate Program (NIH T32HD007470), and Medical Scientist Training Program (NIH T32GM007618). M.P.W. was supported by the AHA Predoctoral and Winslow Fellowship, and M.L. was supported by a Sarnoff Fellowship.

Received: December 17, 2014  
Revised: January 16, 2015  
Accepted: February 18, 2015  
Published: March 12, 2015

## REFERENCES

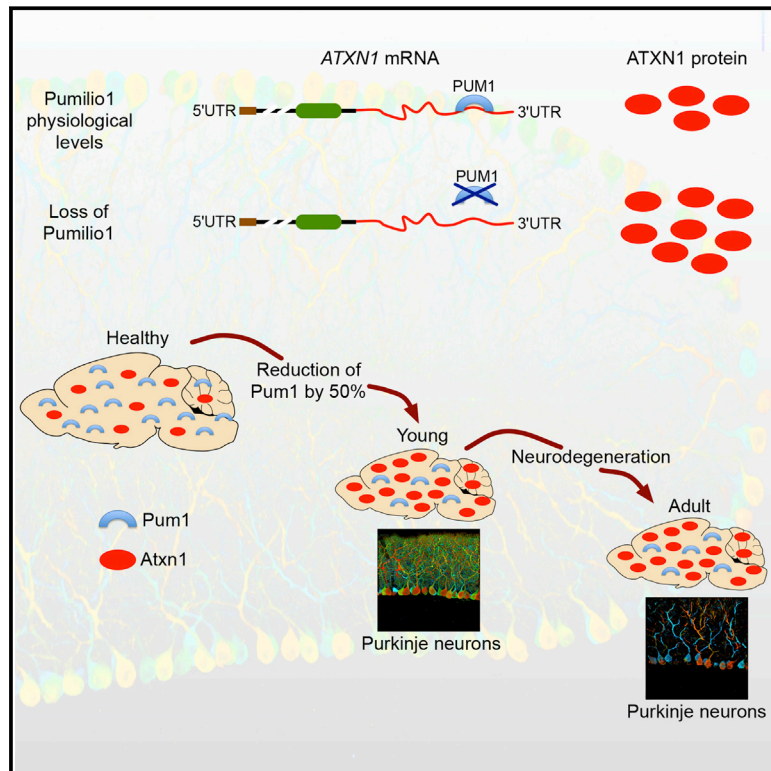
- Alatalo, S.L., Halleen, J.M., Hentunen, T.A., Mönkkönen, J., and Väänänen, H.K. (2000). Rapid screening method for osteoclast differentiation in vitro that measures tartrate-resistant acid phosphatase 5b activity secreted into the culture medium. *Clin. Chem.* 46, 1751–1754.
- Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J., and Nohe, A. (2011). Bone morphogenetic proteins: a critical review. *Cell. Signal.* 23, 609–620.
- Combs, M.D., and Yutzy, K.E. (2009). Heart valve development: regulatory networks in development and disease. *Circ. Res.* 105, 408–421.
- Conway, D.E., Sakurai, Y., Weiss, D., Vega, J.D., Taylor, W.R., Jo, H., Eskin, S.G., Marcus, C.B., and McIntire, L.V. (2009). Expression of CYP1A1 and CYP1B1 in human endothelial cells: regulation by fluid shear stress. *Cardiovasc. Res.* 81, 669–677.
- Deleuze, V., El-Hajj, R., Chalhoub, E., Dohet, C., Pinet, V., Couttet, P., and Mathieu, D. (2012). Angiotensin-2 is a direct transcriptional target of TAL1, LYL1 and LMO2 in endothelial cells. *PLoS ONE* 7, e40484.
- Dennler, S., Prunier, C., Ferrand, N., Gauthier, J.M., and Atfi, A. (2000). c-Jun inhibits transforming growth factor beta-mediated transcription by repressing Smad3 transcriptional activity. *J. Biol. Chem.* 275, 28858–28865.
- Drake, C.J., Brandt, S.J., Trusk, T.C., and Little, C.D. (1997). TAL1/SCL is expressed in endothelial progenitor cells/angioblasts and defines a dorsal-to-ventral gradient of vasculogenesis. *Dev. Biol.* 192, 17–30.
- Engelholm, L.H., Nielsen, B.S., Netzel-Arnett, S., Solberg, H., Chen, X.-D., Lopez Garcia, J.M., Lopez-Otin, C., Young, M.F., Birkedal-Hansen, H., Danø, K., et al. (2001). The urokinase plasminogen activator receptor-associated protein/endo180 is coexpressed with its interaction partners urokinase plasminogen activator receptor and matrix metalloproteinase-13 during osteogenesis. *Lab. Invest.* 81, 1403–1414.
- Foffa, I., Ait Ali, L., Panesi, P., Mariani, M., Festa, P., Botto, N., Vecoli, C., and Andreassi, M.G. (2013). Sequencing of NOTCH1, GATA5, TGFB1 and TGFB2 genes in familial cases of bicuspid aortic valve. *BMC Med. Genet.* 14, 44.
- Garg, V., Muth, A.N., Ransom, J.F., Schluterman, M.K., Barnes, R., King, I.N., Grossfeld, P.D., and Srivastava, D. (2005). Mutations in NOTCH1 cause aortic valve disease. *Nature* 437, 270–274.
- Go, A.S., Mozaffarian, D., Roger, V.L., Benjamin, E.J., Berry, J.D., Blaha, M.J., Dai, S., Ford, E.S., Fox, C.S., Franco, S., et al.; American Heart Association Statistics Committee and Stroke Statistics Subcommittee (2014). Heart disease and stroke statistics—2014 update: a report from the American Heart Association. *Circulation* 129, e28–e292.
- Goettsch, C., Hutcheson, J.D., and Aikawa, E. (2013). MicroRNA in cardiovascular calcification: focus on targets and extracellular vesicle delivery mechanisms. *Circ. Res.* 112, 1073–1084.
- Gorrini, C., Harris, I.S., and Mak, T.W. (2013). Modulation of oxidative stress as an anticancer strategy. *Nat. Rev. Drug Discov.* 12, 931–947.
- Guo, B., Slevin, M., Li, C., Parameshwar, S., Liu, D., Kumar, P., Bernabeu, C., and Kumar, S. (2004). CD105 inhibits transforming growth factor- $\beta$ -Smad3 signalling. *Anticancer Res.* 24 (3a), 1337–1345.
- Hervas-Stubbs, S., Perez-Gracia, J.L., Rouzaut, A., Sanmamed, M.F., Le Bon, A., and Melero, I. (2011). Direct effects of type I interferons on cells of the immune system. *Clin. Cancer Res.* 17, 2619–2627.
- Hofmann, J.J., Briot, A., Enciso, J., Zovein, A.C., Ren, S., Zhang, Z.W., Radtke, F., Simons, M., Wang, Y., and Iruela-Arispe, M.L. (2012). Endothelial deletion of murine Jag1 leads to valve calcification and congenital heart defects associated with Alagille syndrome. *Development* 139, 4449–4460.
- Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41, 178–186.
- Kwa, M.Q., Nguyen, T., Huynh, J., Ramnath, D., De Nardo, D., Lam, P.Y., Reynolds, E.C., Hamilton, J.A., Sweet, M.J., and Scholz, G.M. (2014). Interferon



- regulatory factor 6 differentially regulates Toll-like receptor 2-dependent chemokine gene expression in epithelial cells. *J. Biol. Chem.* 289, 19758–19768.
- Li, D., Chen, X.Q., Li, W.-J., Yang, Y.-H., Wang, J.-Z., and Yu, A.C.H. (2007). Cytoglobin up-regulated by hydrogen peroxide plays a protective role in oxidative stress. *Neurochem. Res.* 32, 1375–1380.
- Li, H., Li, T., Wang, S., Wei, J., Fan, J., Li, J., Han, Q., Liao, L., Shao, C., and Zhao, R.C. (2013). miR-17-5p and miR-106a are involved in the balance between osteogenic and adipogenic differentiation of adipose-derived mesenchymal stem cells. *Stem Cell Res. (Amst.)* 10, 313–324.
- Liao, E.C., Paw, B.H., Oates, A.C., Pratt, S.J., Postlethwait, J.H., and Zon, L.I. (1998). SCL/Tal-1 transcription factor acts downstream of cloche to specify hematopoietic and vascular progenitors in zebrafish. *Genes Dev.* 12, 621–626.
- Luo, G., Ducey, P., McKee, M.D., Pinero, G.J., Loyer, E., Behringer, R.R., and Karsenty, G. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature* 386, 78–81.
- Masumura, T., Yamamoto, K., Shimizu, N., Obi, S., and Ando, J. (2009). Shear stress increases expression of the arterial endothelial marker ephrinB2 in murine ES cells via the VEGF-Notch signaling pathways. *Arterioscler. Thromb. Vasc. Biol.* 29, 2125–2131.
- Mohamed, S.A., Aherrahrou, Z., Liptau, H., Erasmi, A.W., Hagemann, C., Wrobel, S., Borzym, K., Schunkert, H., Sievers, H.H., and Erdmann, J. (2006). Novel missense mutations (p.T596M and p.P1797H) in NOTCH1 in patients with bicuspid aortic valve. *Biochem. Biophys. Res. Commun.* 345, 1460–1465.
- Okuda, T., van Deursen, J., Hiebert, S.W., Grosfeld, G., and Downing, J.R. (1996). AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* 84, 321–330.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.
- Sampson, E.M., Haque, Z.K., Ku, M.C., Tevosian, S.G., Albanese, C., Pestell, R.G., Paulson, K.E., and Yee, A.S. (2001). Negative regulation of the Wnt-beta-catenin pathway by the transcriptional repressor HBP1. *EMBO J.* 20, 4500–4511.
- Schober, A., Nazari-Jahantigh, M., Wei, Y., Bidzhikov, K., Gremse, F., Grommes, J., Megens, R.T.A., Heyll, K., Noels, H., Hristov, M., et al. (2014). MicroRNA-126-5p promotes endothelial proliferation and limits atherosclerosis by suppressing Dlk1. *Nat. Med.* 20, 368–376.
- Shi, X., Richard, J., Zirbes, K.M., Gong, W., Lin, G., Kyba, M., Thomson, J.A., Koyano-Nakagawa, N., and Garry, D.J. (2014). Cooperative interaction of ETV2 and GATA2 regulates the development of endothelial and hematopoietic lineages. *Dev. Biol.* 389, 208–218.
- Smith, N., Dong, Y., Lian, J.B., Pratap, J., Kingsley, P.D., van Wijnen, A.J., Stein, J.L., Schwarz, E.M., O'Keefe, R.J., Stein, G.S., and Drissi, M.H. (2005). Overlapping expression of Runx1(Cbfa2) and Runx2(Cbfa1) transcription factors supports cooperative induction of skeletal development. *J. Cell. Physiol.* 203, 133–143.
- Surapisitchat, J., Jeon, K.I., Yan, C., and Beavo, J.A. (2007). Differential regulation of endothelial cell permeability by cGMP via phosphodiesterases 2 and 3. *Circ. Res.* 101, 811–818.
- Van Handel, B., Montel-Hagen, A., Sasidharan, R., Nakano, H., Ferrari, R., Boogerd, C.J., Schredelseker, J., Wang, Y., Hunter, S., Org, T., et al. (2012). Scl represses cardiomyogenesis in prospective hemogenic endothelium and endocardium. *Cell* 150, 590–605.
- Vu, T.H., and Werb, Z. (2000). Matrix metalloproteinases: effectors of development and normal physiology. *Genes Dev.* 14, 2123–2133.
- Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 151, 206–220.
- Wang, H., Zou, J., Zhao, B., Johannsen, E., Ashworth, T., Wong, H., Pear, W.S., Schug, J., Blacklow, S.C., Arnett, K.L., et al. (2011). Genome-wide analysis reveals conserved and divergent features of Notch1/RBPJ binding in human and murine T-lymphoblastic leukemia cells. *Proc. Natl. Acad. Sci. USA* 108, 14908–14913.
- Weinberg, E.J., Mack, P.J., Schoen, F.J., García-Cardena, G., and Kaazempur Mofrad, M.R. (2010). Hemodynamic environments from opposing sides of human aortic valve leaflets evoke distinct endothelial phenotypes in vitro. *Cardiovasc. Eng.* 10, 5–11.
- White, M.P., Rufaihah, A.J., Liu, L., Ghebremariam, Y.T., Ivey, K.N., Cooke, J.P., and Srivastava, D. (2013). Limited gene expression variation in human embryonic stem cell and induced pluripotent stem cell-derived endothelial cells. *Stem Cells* 31, 92–103.
- Wu, J., Bowe, D.B., Sadlonova, A., Whisenhunt, T.R., Hu, Y., Rustgi, A.K., Nie, Y., Paterson, A.J., and Yang, X. (2014). O-GlcNAc transferase is critical for transducin-like enhancer of split (TLE)-mediated repression of canonical Wnt signaling. *J. Biol. Chem.* 289, 12168–12176.
- Yamashiro, T., Wang, X.-P., Li, Z., Oya, S., Aberg, T., Fukunaga, T., Kamioka, H., Speck, N.A., Takano-Yamamoto, T., and Thesleff, I. (2004). Possible roles of Runx1 and Sox9 in incipient intramembranous ossification. *J. Bone Miner. Res.* 19, 1671–1677.
- Yashiro-Ohtani, Y., Wang, H., Zang, C., Arnett, K.L., Bailis, W., Ho, Y., Knoechel, B., Lanauze, C., Louis, L., Forsyth, K.S., et al. (2014). Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. USA* 111, E4946–E4953.
- Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T.Y., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.

# ***Pumilio1* Haploinsufficiency Leads to SCA1-like Neurodegeneration by Increasing Wild-Type Ataxin1 Levels**

## Graphical Abstract



## Authors

Vincenzo A. Gennarino, Ravi K. Singh, ..., Roy V. Sillitoe, Huda Y. Zoghbi

## Correspondence

hzoghbi@bcm.edu

## In Brief

*Pumilio1* is an RNA-binding protein that binds *Ataxin1* mRNA and regulates its stability. Haploinsufficiency of *Pumilio1* results in an increase in Ataxin1 levels, leading to progressive motor dysfunction and degeneration of Purkinje cells, features typical of spinocerebellar ataxia type 1. These data suggest that either haploinsufficiency of *PUMILIO1* or duplication of *ATAXIN1* could contribute to neurodegeneration in humans.

## Highlights

- The RNA-binding protein PUMILIO1 regulates levels of ATAXIN1 protein and mRNA
- A modest increase in wild-type Ataxin1 levels is enough to cause neurodegeneration
- *Pumilio1* haploinsufficiency accelerates SCA1 disease progression
- *Ataxin1* haploinsufficiency rescues *Pumilio1*<sup>+/-</sup> phenotypes



# ***Pumilio1* Haploinsufficiency Leads to SCA1-like Neurodegeneration by Increasing Wild-Type Ataxin1 Levels**

Vincenzo A. Gennarino,<sup>1,10</sup> Ravi K. Singh,<sup>2</sup> Joshua J. White,<sup>2,3,10</sup> Antonia De Maio,<sup>4,10</sup> Kihoon Han,<sup>1,5,10</sup> Ji-Yoen Kim,<sup>1,10</sup> Paymaan Jafar-Nejad,<sup>1,10,12</sup> Alberto di Ronza,<sup>1,10</sup> Hyojin Kang,<sup>1,10,13</sup> Layal S. Sayegh,<sup>1,10,14</sup> Thomas A. Cooper,<sup>2,4,6,7,8</sup> Harry T. Orr,<sup>11</sup> Roy V. Sillitoe,<sup>2,3,4,10</sup> and Huda Y. Zoghbi<sup>1,2,4,5,9,10,\*</sup>

<sup>1</sup>Department of Molecular and Human Genetics

<sup>2</sup>Department of Pathology and Immunology

<sup>3</sup>Department of Neuroscience

<sup>4</sup>Program in Developmental Biology

<sup>5</sup>Howard Hughes Medical Institute

<sup>6</sup>Department of Molecular and Cellular Biology

<sup>7</sup>Dan L. Duncan Cancer Center

<sup>8</sup>Department of Molecular Physiology and Biophysics

<sup>9</sup>Department of Pediatrics

Baylor College of Medicine, Houston, TX 77030, USA

<sup>10</sup>Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX 77030, USA

<sup>11</sup>Institute for Translational Neuroscience, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA

<sup>12</sup>Present address: Isis Pharmaceuticals, Inc. 2855 Gazelle Court, Carlsbad, CA 92010, USA

<sup>13</sup>Present address: HPC-enabled Convergence Technology Research Division, Korea Institute of Science and Technology Information, Daejeon, South Korea

<sup>14</sup>Present address: Emory University School of Medicine, Atlanta, GA 30322, USA

\*Correspondence: [hzoghibi@bcm.edu](mailto:hzoghibi@bcm.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.012>

## **SUMMARY**

Spinocerebellar ataxia type 1 (SCA1) is a paradigmatic neurodegenerative proteinopathy, in which a mutant protein (in this case, ATAXIN1) accumulates in neurons and exerts toxicity; in SCA1, this process causes progressive deterioration of motor coordination. Seeking to understand how post-translational modification of ATAXIN1 levels influences disease, we discovered that the RNA-binding protein PUMILIO1 (PUM1) not only directly regulates ATAXIN1 but also plays an unexpectedly important role in neuronal function. Loss of *Pum1* caused progressive motor dysfunction and SCA1-like neurodegeneration with motor impairment, primarily by increasing Ataxin1 levels. Breeding *Pum1*<sup>+/-</sup> mice to SCA1 mice (*Atxn1*<sup>154Q/+</sup>) exacerbated disease progression, whereas breeding them to *Atxn1*<sup>+/-</sup> mice normalized Ataxin1 levels and largely rescued the *Pum1*<sup>+/-</sup> phenotype. Thus, both increased wild-type ATAXIN1 levels and *PUM1* haploinsufficiency could contribute to human neurodegeneration. These results demonstrate the importance of studying post-transcriptional regulation of disease-driving proteins to reveal factors underlying neurodegenerative disease.

## **INTRODUCTION**

Misfolded proteins underlie the pathogenesis of a number of neurodegenerative conditions, collectively known as proteinopathies. Alzheimer disease (AD), Parkinson disease (PD), amyotrophic lateral sclerosis (ALS), and polyglutamine diseases such as Huntington disease all fall into this category (Ross and Poirier, 2004; Soto, 2003). Despite the heterogeneity of their pathogenic mechanisms, in each of these diseases, the misfolded protein accumulates in neurons and exerts toxicity. Somewhat surprisingly, the brain can also be sensitive to elevated levels of wild-type (WT) protein: duplication of the *amyloid precursor protein* (APP) locus causes autosomal dominant early-onset AD (Rovelet-Lecrux et al., 2006; Rumble et al., 1989), and duplications or triplications of *α-synuclein* (SNCA) are associated with familial PD (Chartier-Harlin et al., 2004; Ibáñez et al., 2004; Singleton et al., 2003). Along similar lines, it has been shown recently that *leucine-rich repeat kinase 2* (LRRK2) mutations, the most common cause of inherited PD, increase overall protein synthesis in *Drosophila*, and that reduction in dLRRK levels is protective (Martin et al., 2014).

Spinocerebellar ataxia type 1 (SCA1) is paradigmatic of the subgroup of polyglutamine (polyQ) proteinopathies caused by expansion of an unstable CAG repeat in the coding region of the relevant disease gene, in this case *ATAXIN1* (*ATXN1*) (Orr et al., 1993). The onset of SCA1 is usually in mid-life, when motor coordination begins to deteriorate because of cerebellar degeneration; patients eventually die of bulbar dysfunction that

renders them unable to clear their airway (Zoghbi and Orr, 2009). There is clear evidence that the expanded polyQ tract stabilizes ATXN1 and causes it to resist being cleared by the ubiquitin-proteasome pathway, in effect increasing its abundance in neurons (Cummings et al., 1999). Notably, the severity of neurodegeneration in fly and mouse models of SCA1 correlates directly with levels of mutant ATXN1 protein (Burrigh et al., 1995; Fernandez-Funez et al., 2000), and massive overexpression of even WT ATXN1 under the Purkinje-cell-specific promoter can produce a mild SCA1-like phenotype in mice (Fernandez-Funez et al., 2000).

Although the artificiality of transgenic models limits their relevance to the human disease, these results from SCA1 transgenic mice, along with the evidence from familial AD and PD patients, led us to ask whether there were post-transcriptional modifications that might increase the levels of WT ATXN1 in a more physiologically relevant way and shed further light on the role of protein levels in neurodegeneration. The extraordinarily long 3' UTR, approximately 7 kb in *ATXN1* mRNA, seemed to promise a rich source of key brain-enriched post-transcriptional regulatory elements. To our surprise, we found that ATXN1 is regulated directly by an RNA-binding protein (RBP), *Pumilio1*, and that a brain-wide increase in WT *Atxn1* levels of only ~50%, caused by *Pum1* haploinsufficiency, is sufficient to cause marked neurodegeneration in mice.

## RESULTS

### The RBP PUMILIO1 Regulates ATAXIN1 Levels in Cells

Two types of molecules are known to modulate protein levels by binding to the corresponding mRNA: RBPs and microRNAs (miRNAs). RBPs bind to specific sequence motifs or secondary structures in mRNAs and regulate multiple steps in RNA metabolism, such as splicing, nucleus-cytoplasm transport, and translation (Lukong et al., 2008). On the other hand, miRNAs are small non-coding RNAs that control various developmental and physiological processes by suppressing the expression of their target genes via binding of a short (6–8 nucleotide) complementary seed region in the 3' UTRs of mRNAs (Bartel, 2009).

We first scanned the ~7 kb-long *ATXN1* 3' UTR for potential binding sites for miRNAs by using the TargetScan (Friedman et al., 2009), CoMeTa (Gennarino et al., 2012), and HOCTARdb (Gennarino et al., 2011) prediction tools. As expected, scanning identified dozens of potential miRNA-binding sites (data not shown). Because RNA folding mediates miRNA-RNA interactions by masking or exposing specific binding-site sequences, we analyzed the secondary structure of the *ATXN1*-3' UTR (Wan et al., 2014) to prioritize the best candidate *ATXN1*-modulating miRNAs. This revealed a complicated secondary structure that masks the binding sites for almost all of the putative miRNAs that might target the *ATXN1*-3' UTR (Figure S1). For miRNAs to act on *ATXN1* mRNA, they would likely require the help of RBPs to unfold such a structure.

Scan analysis of the human *ATXN1*-3' UTR revealed three putative *Pumilio1* (PUM1) binding motifs (Wang et al., 2002) at positions 682, 2812, and 5275 from the beginning of the UTR (Figure 1A). The RBP PUM1 regulates its target genes by inducing a conformational switch in the 3' UTR that unmasks

specific miRNA-binding sites (Kedde et al., 2010; Miles et al., 2012). Interestingly, the motif in position 5275 (Figure 1A, red box) is highly conserved across several species and represents the canonical PUM1-binding motif (5'-UGUAXUA-3') (Galgano et al., 2008; Wang et al., 2002). Overexpressing *PUM1* in HEK293T cells reduced *ATXN1* mRNA levels, whereas decreasing *PUM1* by two different RNAi increased *ATXN1* mRNA levels (Figure 1B). In vitro overexpression of *PUM1* consistently decreased the luciferase activity of a reporter construct expressing the full-length *ATXN1*-3' UTR (Figure 1C). Mutation of each PUM1-binding motif within the *ATXN1*-3' UTR revealed that only the most conserved site, containing the canonical motif, is functional; when mutated, it abolished the effect of *PUM1* overexpression on luciferase activity (Figure 1D).

### Pum1 Is Widely Expressed in Mouse Brain and Regulates *Atxn1* Levels In Vivo

To examine the endogenous expression pattern of *Pum1* in mice, we performed in situ hybridization assays (ISH) and western blot on 3-week-old mouse brain sections. *Pum1* was expressed in all major brain regions in WT mice, almost completely absent in the brain of null mice, and reduced in heterozygous (*Pum1*<sup>+/-</sup>) brains (Figure S2A). We also confirmed that *Pum1* protein is widely expressed in the brain at 5 weeks of age (Figure S2B).

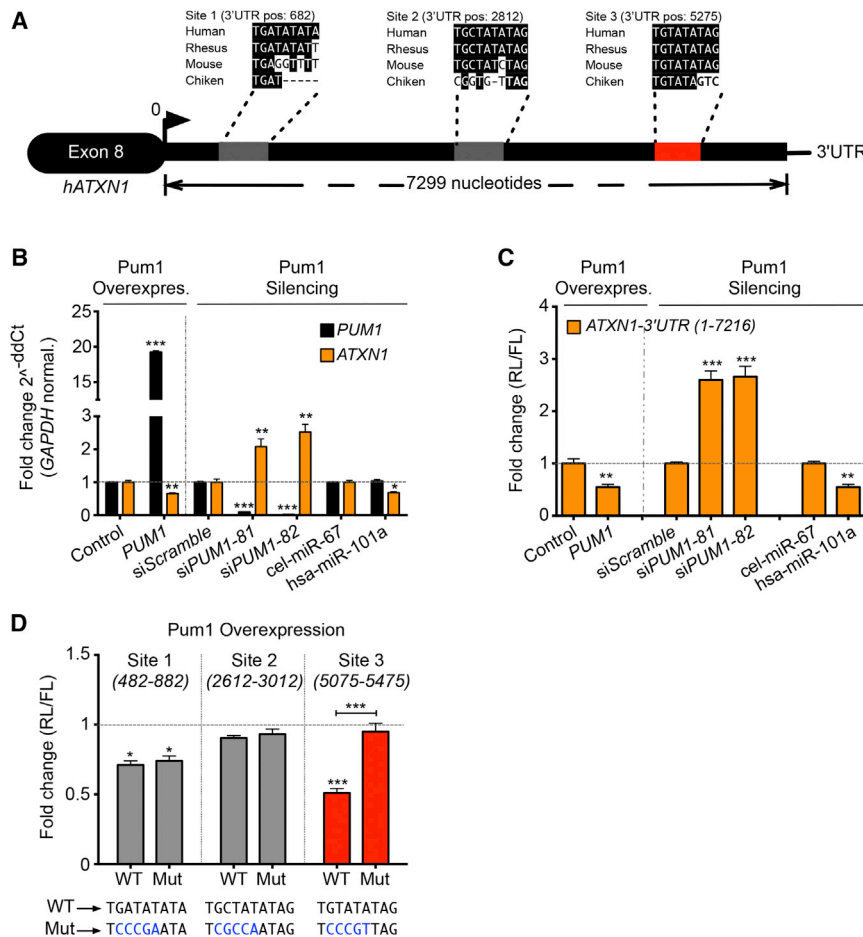
To determine whether *Pum1* binds *Atxn1* mRNA in vivo, we performed an RNA cross-linking and immunoprecipitation assay (RNA-Clip) on cerebra and cerebella from 5-week-old WT animals, using *Pum1* knockout mice (*Pum1*<sup>-/-</sup>) as negative controls (Figure S2C). We found that *Pum1* physically interacts with the conserved binding site of the *Atxn1*-3' UTR in WT mice (Figure 2A). Consistent with the finding that *Pum1* negatively regulates *Atxn1*, *Pum1* heterozygous (*Pum1*<sup>+/-</sup>) mice showed increased levels of both *Atxn1* protein (Figure 2B) and mRNA (Figure 2C)—approximately 30% in the cerebrum and 50% in the cerebellum—and *Pum1*<sup>-/-</sup> mice showed even more pronounced increases (Figures 2B and 2C). These data demonstrate that *Pum1* directly regulates *Atxn1* levels in the mouse brain.

### PUM1 Controls ATXN1 Levels by Affecting RNA Stability and not through the miRNA Machinery

Several mRNA subsets contain target sites for both RBPs and miRNAs, and cooperation between these two types of post-transcriptional regulators has been described (Bhattacharyya et al., 2006; Fabian and Sonenberg, 2012; Glorian et al., 2011; Kim et al., 2009; Kundu et al., 2012). This may be particularly relevant for PUM1, as studies have indicated extensive interaction between PUM1 and the miRNA regulatory system (Kedde et al., 2010; Galgano et al., 2008).

To determine whether PUM1 regulates ATXN1 through miRNA by inducing a conformational switch in its 3' UTR, we overexpressed *PUM1* in HEK293T cells along with miR-101a, a miRNA known to modulate *ATXN1* levels (Lee et al., 2008). These conditions significantly reduced levels of ATXN1 protein (Figures 3A and S3A) and mRNA (Figure S3E), but no more than overexpressing miR-101a or *PUM1* separately. In fact, overexpression





**Figure 1. PUM1 Regulates ATXN1 Levels via a Highly Conserved Binding Motif**

(A) Schematic representation of human *ATXN1*-3' UTR showing three putative PUM1-binding motifs (gray and red boxes) and their conservation in different species. The numbers indicate positions of PUM1 motifs in the human *ATXN1*-3' UTR. (B) *ATXN1* mRNA quantification by qRT-PCR in HEK293T cells upon overexpression (left panel) or knockdown (siPUM1-81 and -82) (right panel) of *PUM1*. The destination-cloning vector (control), scrambled siRNAs (siScr.), and cel-miR-67 were used as negative controls. The housekeeping gene *GAPDH* was used to normalize the expression of genes in all the qRT-PCR experiments. (C) Luciferase assay in HEK293T cells overexpressing the reporter construct harboring the full-length *ATXN1*-3' UTR. In these conditions, we overexpressed (left panel) or decreased expression (right panel) of *PUM1*.

(D) Luciferase assay in HEK293T cells transfected with single WT and mutant (MUT) putative PUM1-binding site on the *ATXN1*-3' UTR. The positions of cloned regions for each PUM1-binding motif are indicated. The mutagenized nucleotides are highlighted in blue.

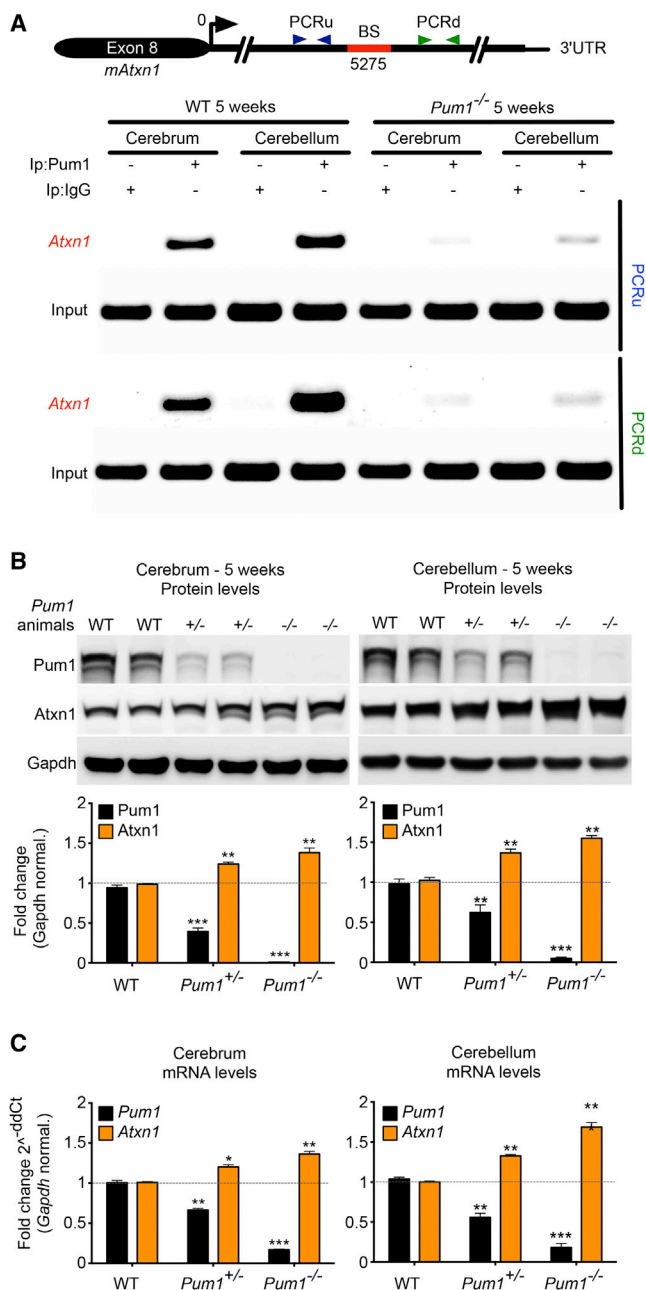
In (C) and (D), the destination-vector (control), RNAi scramble (siScramble), and cel-miR-67 were used as negative controls; miR-101a was used as positive control. (RL) Renilla, (FL) Firefly luciferase. All the experiments were performed in triplicate (data represent mean  $\pm$  SEM). *p* values were calculated by Student's *t* test. Statistical significance is indicated as follows: \**p* < 0.05, \*\**p* < 0.01, \*\*\**p* < 0.0001. See also Figure S1.

of miR-101a along with *PUM1* knockdown consistently decreased levels of *ATXN1* protein (Figures 3B and S3B) and mRNA (Figure S3F) to a degree comparable to that of miR-101a overexpression alone. These results suggest that *PUM1* regulates *ATXN1* in a miR-101a-independent fashion but do not exclude the possibility that other miRNAs bind the *ATXN1* 3' UTR. To obviate testing the effect of *PUM1* on all possible miRNAs regulating *ATXN1*, we knocked down the catalytic engine of the RNA-induced silencing complex (RISC), *Argonaute-2* (*AGO2*), to globally inhibit miRNA binding and retested *PUM1*'s ability to regulate *ATXN1*. We found that *PUM1* overexpression in the context of *AGO2* knockdown still reduced levels of both *ATXN1* protein (Figures 3C and S3C) and mRNA (Figure S3G). Conversely, simultaneous RNAi of *PUM1* and *AGO2* increased levels of both *ATXN1* protein (Figures 3D and S3D) and mRNA (Figure S3H), but no more than silencing *PUM1* alone. These data establish that *PUM1* modulates *ATXN1* levels directly by binding its 3' UTR, without the assistance of the miRNA machinery.

To further explore the mechanism by which *PUM1* regulates *ATXN1* levels, we tested whether *PUM1* influences the stability or the translation of *ATXN1* mRNA. We transfected HEK293T with a luciferase reporter encoding an *ATXN1*-3' UTR harboring either the conserved WT or mutated (Mut) *PUM1*-binding site.

Later, we used treatment with DRB (5,6-dichloro-1- $\beta$ -D-ribofuranosylbenzimidazole), a drug that inhibits RNA translation by blocking RNA polymerase II in the early elongation stage, to assess the levels of the reporter transcript. Upon the addition of DRB (time-point zero), the relative expression of reporter transcripts containing the *ATXN1*-3' UTR Mut binding site is considerably higher than that of transcripts containing the *ATXN1*-3' UTR WT binding site (Figure 3E, top panel). This difference remains stable over time until 8 hr after DRB addition. Remarkably, the *ATXN1*-3' UTR with Mut binding site reached its half-life after 19 hr, whereas the *ATXN1*-3' UTR with WT binding site decreased linearly over time, reaching its half-life at nearly 8 hr (Figure 3E, top panel). Given that the promoter sequences of the *ATXN1*-3' UTR constructs carrying either WT or Mut binding sites are exactly the same and that transfection of neither construct affected *PUM1* protein levels, we conclude that *PUM1* promotes degradation of *ATXN1* by binding its 3' UTR (Figures 3E, bottom panel and S3I).

To investigate physiological changes in *ATXN1* mRNA, we decided to knock down *PUM1* in HEK293T cells and measure the half-life of endogenous *ATXN1* mRNA at different time points after DRB treatment. Knockdown of *PUM1* (siPUM1) was associated with a significant increase of *ATXN1* mRNA from time zero and remained upregulated up to 8 hr after translation



**Figure 2. Pum1 Directly Binds the 3' UTR of *Atxn1* to Regulate Its Levels in Mouse Cerebrum and Cerebellum**

(A) RNA-clip for the conserved Pum1-binding site in mouse cerebrum and cerebellum. PCRu and PCRD highlight the PCR fragments upstream and downstream of the conserved *Atxn1*-3' UTR Pum1-binding site (BS). IP with IgG as well as *Pum1* null mice were used as negative controls. Isolated RNA from a fraction (10%) of pre-cleared lysate was used as input. The experiment was performed in triplicate.

(B and C) Quantification of *Atxn1* protein (B) and mRNA levels (C) in WT and *Pum1*<sup>-/-</sup> mice in cerebrum and cerebellum (n = 8 per genotype).

Data represent mean ± SEM and normalized to Gapdh. See [Experimental Procedures](#) for more details. p values were calculated by Student's t test; \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.0001. See also [Figure S2](#).

inhibition (Figure 3F, top panel). Our calculation consistently showed that the half-life of *ATXN1* mRNA was much longer (nearly 12 hr) after siPUM1 than after siScramble transfection (~4 hr) (Figure 3F, top panel). We confirmed PUM1 downregulation by quantifying mRNA at time zero (Figure S3J) and protein levels at different time points (Figures 3F, bottom panel and S3K). PUM1 thus increases *ATXN1* levels by directly regulating the stability of *ATXN1* mRNA.

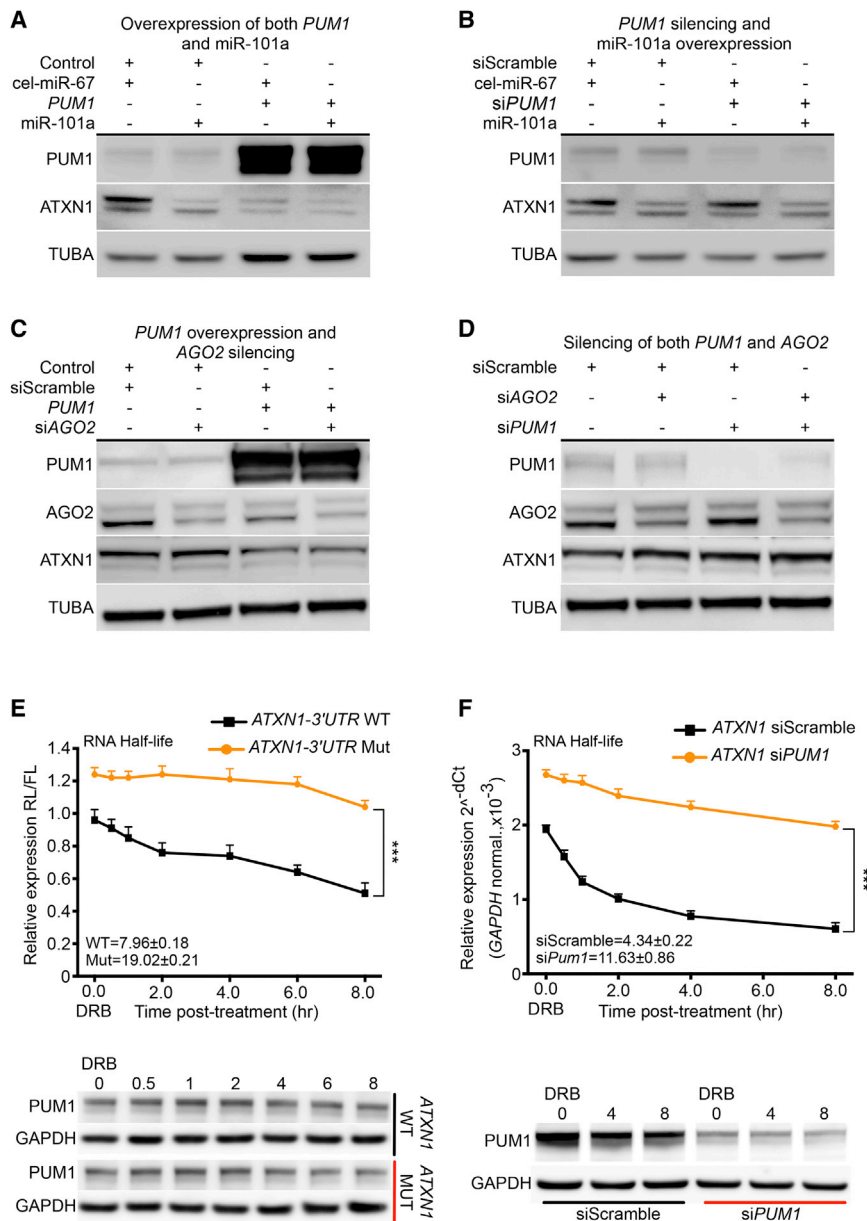
### Pum1 Mutant Mice Develop Progressive Motor Dysfunction and Neurodegeneration

Recent studies have shown that Pum1 is an essential regulator of spermatogenesis in mice and promotes differentiation of embryonic stem cells (Chen et al., 2012; Leeb et al., 2014), but its role in the mammalian nervous system has not been investigated. We therefore characterized the brain structure and behavior of *Pum1* knockout mice (Chen et al., 2012).

The *Pum1* null allele tends to be transmitted with an altered Mendelian ratio (Figure S4A). Compared to WT and *Pum1*<sup>+/-</sup> littermates, *Pum1*<sup>-/-</sup> mice were significantly smaller in body length, weight, and brain weight and size (Figures 4A and S4B). Surprisingly, the loss of one copy of *Pum1* was sufficient to cause impaired performance on the accelerating rotarod assay in 5-week-old mice (Figure 4B): the motor deficit had progressed in severity by 12 weeks (Figure S4C). This motor incoordination was even more dramatic in *Pum1*<sup>-/-</sup> age-matched mice (Figures 4B and S4C), which performed equally poorly in the dowel-walking test (Figures S4D and S4E)—as poorly, in fact, as SCA1 mice at this age in both assays (Watake et al., 2002).

Beginning at 8 weeks of age, both *Pum1*<sup>+/-</sup> and *Pum1*<sup>-/-</sup> mice exhibited hind-paw claspings when suspended by the tail (Figures 4C and S4F), a sign of neurological dysfunction. At 10 weeks of age, *Pum1*<sup>-/-</sup> mice displayed significantly less vertical activity in an open-field chamber (Figure S4G) and spent less time in its center (Figure S4H) but traveled greater distances over 30 min in the chamber than WT (Figure 4D). Interestingly, *Pum1*<sup>-/-</sup> mice covered a greater distance at 18 weeks than at 10 weeks of age (Figure S4I). Using the DigiGait assay, we found that 12-week-old *Pum1*<sup>-/-</sup> mice had wider stances (Figure S4J), shorter stride lengths (Figure S4K), and greater stride frequencies (Figure S4L). At the same age, *Pum1*<sup>+/-</sup> and *Pum1*<sup>-/-</sup> mice were poor nest builders (Figure S4M). *Pum1* deficiency thus causes progressive loss of motor coordination that appears to be cerebellar in origin.

To uncover the defects underlying the phenotype, we performed neuropathological studies. At 3 and 4 weeks of age, there was no evidence of Purkinje cell pathology in *Pum1*<sup>+/-</sup> or *Pum1*<sup>-/-</sup> mice (Figures 4F and S4N), but by 10 weeks, *Pum1* haploinsufficiency had caused loss of Purkinje cells (Figures 4E and 4F) and dendritic arborization (Figure 4G). Both defects were more dramatic in age-matched *Pum1*<sup>-/-</sup> mice (Figures 4E–4G). The neuronal loss is thus a result of neurodegeneration and not a developmental defect. Notably, progressive Purkinje cell degeneration and motor dysfunction are hallmarks of SCA1 in both human patients and the SCA1 knockin mouse model (*Atxn1*<sup>154Q/+</sup>) (Watake et al., 2002). Because the progressive defects in *Pum1* mutant mice were reminiscent of those



**Figure 3. PUM1 Modulates the Levels of WT ATXN1 Independently of miRNAs**

(A–D) Representative western blot (upper panel) of protein lysates from HEK293T cells upon (A) overexpression of both *PUM1* and miR-101a; (B) RNAi *PUM1* (si*PUM1*) followed by overexpression of miR-101a; (C) overexpression of *PUM1* followed by RNAi AGO2 (siAGO2); and (D) RNAi of both *PUM1* and AGO2. The negative controls were destination-cloning vector (control), RNAi scramble (siScramble), and cel-miR-67. All data were normalized to  $\alpha$ -tubulin (TUBA).

(E) mRNA half-life quantification of WT and Mut *PUM1* ATXN1-3' UTR binding sites in HEK293T cells at different time points upon DRB treatment (time zero). The numeric values within the panel given the extrapolated half-life for WT and Mut RNA,  $p = 3.6 \times 10^{-6}$ . Firefly (FL) RNA levels were quantified and normalized to Renilla (RL). Bottom panel: representative western blot of *PUM1* in HEK293T cells at different time points. Data were normalized to GAPDH.

(F) ATXN1 mRNA half-life quantification in HEK293T cells at different time points, from zero (DRB treatment) to 8 hr total upon RNAi of *PUM1* (si*PUM1*) or RNAi of scramble (siScramble) transfection. The numeric values within the panel given the extrapolated half-life for si*PUM1* and siScramble RNA,  $p = 0.012$ . Bottom panel: representative western blot of *PUM1* in HEK293T cells at different time points.

Data were normalized to GAPDH mRNA (top panel) or protein (bottom panel). All experiments were performed in triplicate (data represent mean  $\pm$  SEM); \*\*\* $p < 0.0001$ . See also Figure S3.

observed in SCA1 mice (Watase et al., 2002), we decided to dissect the genetic interaction between *Pum1* and SCA1 mice.

#### **Pum1 Haploinsufficiency Worsens the Phenotype of the SCA1 Knockin Mouse**

Given that *Atxn1* levels were increased in *Pum1*<sup>+/-</sup> and in *Pum1*<sup>-/-</sup> mice (Figure 2B), we predicted that halving the dosage of *Pum1* in the SCA1 knockin mice (*Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup>) would exacerbate the SCA1 phenotype, and this proved to be the case. *Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> mice tended to be smaller than *Atxn1*<sup>154Q/+</sup> and the other genotypes (lower body and brain weights, shorter lengths, smaller brain sizes; Figures S5A and S5B) and showed more severe motor incoordination and hind-paw claspings than either *Atxn1*<sup>154Q/+</sup> or *Pum1*<sup>+/-</sup> mice (Figures 5A and 5B). Notably,

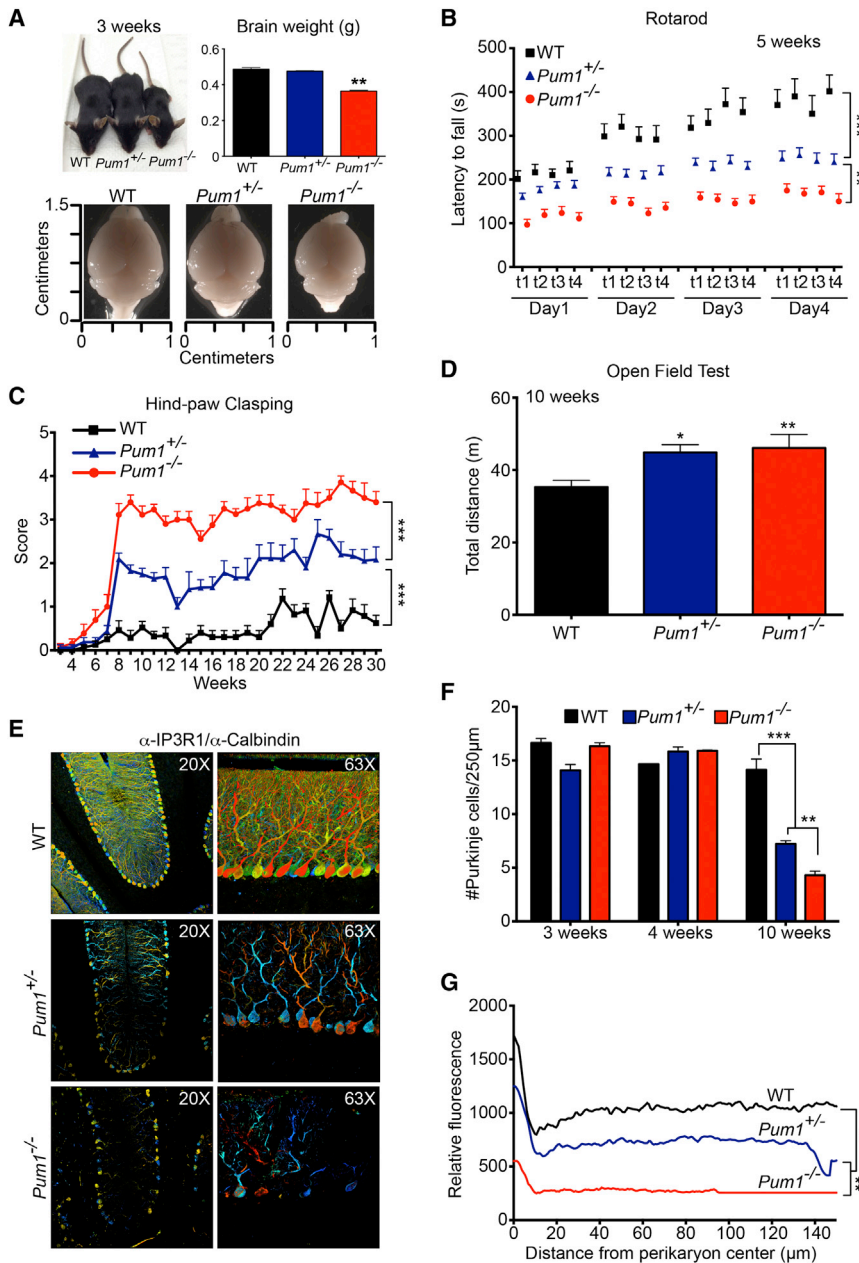
*Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> mice began to show the hind-paw claspings phenotype at 6 weeks, much earlier than age-matched *Atxn1*<sup>154Q/+</sup> mice or *Pum1* single mutants (Figure 5B). At 10 weeks of age, *Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> mice traveled greater distances than WT and *Atxn1*<sup>154Q/+</sup> but not more than *Pum1*<sup>+/-</sup> mice (Figure S5C). Severe kyphosis (curvature of the spine) developed in *Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> mice

8 weeks earlier than SCA1 knockin mice (20 versus 28 weeks; Figure 5C), confirming the accelerated disease course. In addition, *Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> mice had a significantly shorter lifespan than their *Atxn1*<sup>154Q/+</sup> littermates (Figure 5D). At 12 weeks of age, the Purkinje cell loss (Figures 5E and 5F) and arborization defects (Figure 5G) were more dramatic in *Pum1*<sup>+/-</sup>; *Atxn1*<sup>154Q/+</sup> than with all other genotypes. These results suggest a genetic interaction between *Pum1* and *Atxn1*<sup>154Q</sup>.

#### **Genetic Reduction of Atxn1 Levels Rescued the Pum1 Mutant Phenotype**

To test our hypothesis that the neurological deficits of *Pum1* mutant mice resulted from an increase in *Atxn1* levels due to loss of *Pum1* regulation, we crossed *Pum1*<sup>+/-</sup> with *Atxn1*<sup>+/-</sup>





**Figure 4. *Pum1* Mutant Mice Develop Progressive Motor Deficits and Cerebellar Degeneration**

(A) Representative pictures of 3-week-old mice. Body size, brain weight, and brain size are reduced in  $Pum1^{-/-}$  animals.

(B) Accelerating rotarod analysis. Mice were trained over 4 days with four trials (t) per day. The null mice were significantly different from WT from day 1; by day 2, the difference between WT and both  $Pum1$  null and heterozygotes was statistically significant, as was the difference between the two mutants.

(C) Hind-paw clamping analysis in mice: a higher score indicates a more severe phenotype (see Figure S4F, bottom panel for scoring details). By 6 weeks of age, the null mice were statistically different from WT; by 8 weeks, both mutant lines were statistically significantly different from WT.

(D) Open-field test measuring the total distance traveled of the  $Pum1$  null mice relative to WT.

(E) Representative images of immunofluorescence (IF) confocal microscopy in 3D depth-coding (see Experimental Procedures). Co-staining with  $\alpha$ -IP3R1 and -calbindin antibodies was used to label Purkinje cells and to reveal their arborization.

(F) Purkinje cell counts at 3, 4, and 10 weeks old for all examined genotypes.

(G) IF for Calbindin and IP3R1 were quantified and averaged in selected rectangular cerebellar sub-sections. All experiments were performed in WT,  $Pum1^{+/-}$ , and  $Pum1^{-/-}$  mice. More than 12 mice per genotype were considered in (A)–(D) and 6 per genotype in (E)–(G). Data in (A), (D), and (F) represent mean  $\pm$  SEM; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ . See also Figure S4.

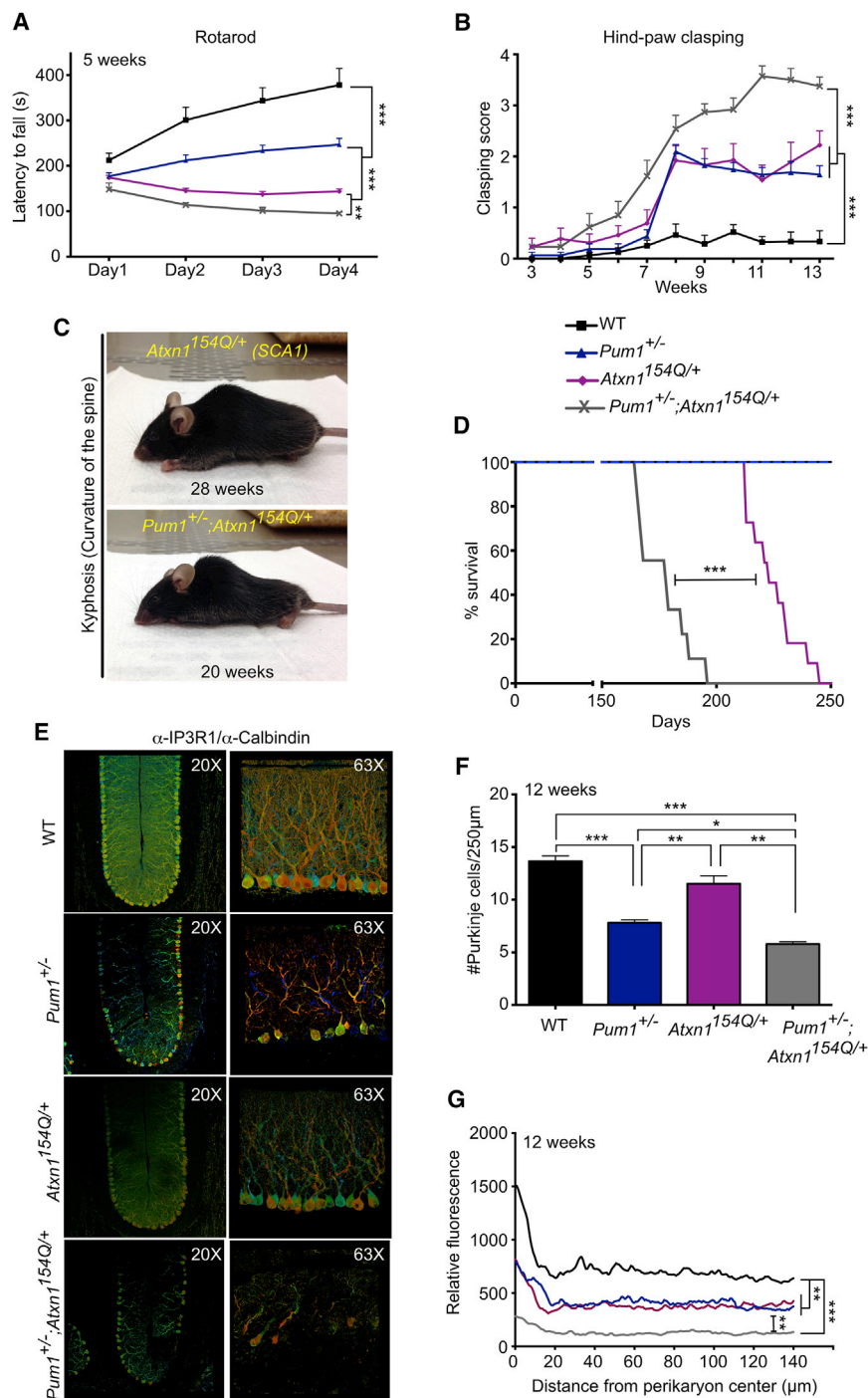
mice and characterized the offspring. We first confirmed that  $Pum1$  haploinsufficiency in the  $Atxn1^{+/-}$  mice ( $Pum1^{+/-}$ ;  $Atxn1^{+/-}$ ) at 5 weeks completely rescued the physiological protein levels of Atxn1 (Figures 6A, S6A, and S6B). At 9 weeks,  $Pum1^{+/-}$ ;  $Atxn1^{+/-}$  mice showed no difference in body weight (Figure S5A), length, brain weight, or brain size relative to any other genotype (Figure S6C).  $Atxn1$  haploinsufficiency in  $Pum1$  mutant mice significantly mitigated the motor deficits observed in  $Pum1^{+/-}$  animals at 5 weeks of age (Figure 6B) and completely rescued the hind-paw clamping (Figure 6C) and kyphosis, which occurs at a later age of  $\sim 25$  weeks (Figure 6D). Interestingly,  $Pum1^{+/-}$ ;  $Atxn1^{+/-}$  mice still traveled farther than other geno-

types and not less than  $Pum1^{+/-}$  mice (Figure S6D). The Purkinje cell loss (Figures 6E and 6F) and arborization defects (Figure 6G) typical of 10-week-old  $Pum1^{+/-}$  mice were rescued in  $Pum1^{+/-}$ ;  $Atxn1^{+/-}$  cerebella. This indicates that, whatever the other pathways affected by  $Pum1$  deficiency, the cerebellar degeneration, clamping, and kyphosis are caused primarily by the influence of  $Pum1$  on Atxn1 levels.

## DISCUSSION

Accumulation of mutant proteins in the brain has been known for some time to underlie the progression of neurodegenerative





**Figure 5. *Pum1* Haploinsufficiency Exacerbates SCA1 Disease Progression**

(A and B) A 50% reduction of *Pum1* in  $Atxn1^{154Q/+}$  mice aggravates (A) motor incoordination on the accelerating rotarod at 5 weeks ( $n = 12$  per genotype) and (B) hind-paw claspings when mice are suspended by the tail ( $n > 12$  mice per genotype). The double mutants were statistically different from WT beginning at week 6. See [Experimental Procedures](#) for details.

(C) Representative picture of SCA1 and  $Pum1^{+/-};Atxn1^{154Q/+}$  mice.  $Pum1^{+/-};Atxn1^{154Q/+}$  showed severe kyphosis (curvature of the spine) at earlier stages than their SCA1 counterparts.

(D) Haploinsufficiency of *Pum1* reduces lifespan in SCA1 background mice. p value was calculated by long-rank test; \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ .

(E) Representative images of IF confocal microscopy in 3D depth-coding (see [Experimental Procedures](#)). Co-staining with  $\alpha$ -IP3R1 and -calbindin antibodies was used to label Purkinje cells and to reveal their arborization.

(F) Purkinje cell count at 12 weeks for all examined genotypes.

(G) IF for calbindin and IP3R1 were quantified and averaged in selected rectangular cerebellar subsections.

Four mice per genotype in (E), (F), and (G) were considered. Data in (F) and (G) represent mean  $\pm$  SEM; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ . See also [Figure S5](#).

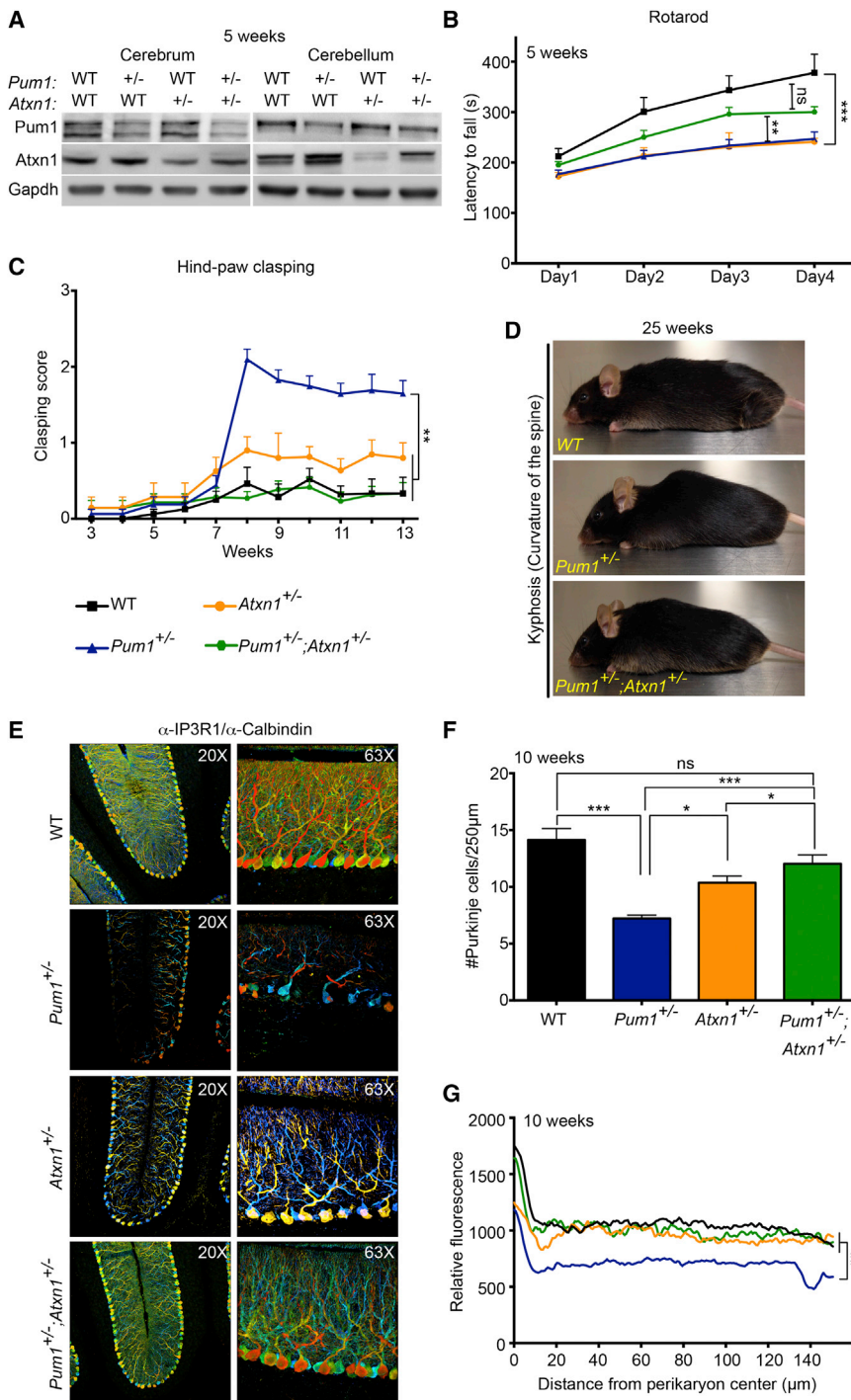
disease course, however, and clearly result from processes that have been taking place for decades. Here we asked how protein levels affect the brain long before aggregates form. Specifically, we sought post-transcriptional regulatory mechanisms that regulate WT ATXN1 levels independently of the polyglutamine tract in order to determine whether reducing protein levels might delay disease progression. Our studies have unexpectedly revealed two candidate genes for neurodegenerative conditions in humans: WT (but upregulated) *ATAXIN1* and *PUMILIO1*.

We began our investigation by scanning the long 3' UTR of ATXN1 to identify regulatory elements that could be used to modulate ATXN1 levels. We found three PUM1-binding motifs in the 3' UTR

disorders such as AD, PD, HD, ALS, and the spinocerebellar ataxias. In all these diseases, the mutant proteins form insoluble aggregates (Haass and Selkoe, 2007; Klement et al., 1998; Ross and Poirier, 2004; Zoghbi and Orr, 2000). Considerable attention has been devoted to the question of whether inhibiting aggregate formation or promoting their dissolution would mitigate disease (Arrasate et al., 2004; Bowman et al., 2005, 2007). The aggregates in all these proteinopathies appear late in the

of ATXN1, one of which is highly conserved. We used mutagenesis and RNA-Clip to show that *Pum1* regulates *Atxn1* levels by binding directly to the highly conserved motif in its 3' UTR.

*Pum1* is a member of a well-characterized family of RBPs, known as the PUF family, which are involved in various physiological processes (Spasov and Jurecic, 2003; Wickens et al., 2002). A typical feature of these proteins is the presence of an RNA-binding *Pumilio* homology domain (PUM-HD) that binds a



**Figure 6. *Atxn1* Haploinsufficiency Rescues Motor Deficits and Cerebellar Pathology in *Pum1*<sup>+/-</sup> Mice**

(A) Representative western blot showing that haploinsufficiency of *Pum1* restores physiological *Atxn1* protein levels. All experiments were performed in triplicate in cerebra and cerebella from mice at 5 weeks of age (data represent mean  $\pm$  SD). All data were normalized to Gapdh. (B and C) *Pum1*<sup>+/-</sup>; *Atxn1*<sup>+/-</sup> mice showed (B) significant improvement in motor performance on the accelerating rotarod ( $n = 12$  per genotype), (C) reduced hind-paw clasping ( $n > 12$  per genotype), and (D) reduced kyphosis (curvature of the spine; photo taken at 25 weeks). (D–G) Purkinje cells loss (E and F) and loss of dendritic arborization (G) were rescued in *Pum1*<sup>+/-</sup>; *Atxn1*<sup>+/-</sup> mice ( $n = 6$  per genotype). Staining was performed with calbindin/IP3R1 in 3D depth-coding images. Data represent mean  $\pm$  SEM. See [Experimental Procedures](#). p values were calculated by Student's t test. ns = not significant; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.0001$ . See also [Figure S6](#).

miRNAs (Fabian and Sonenberg, 2012; Friend et al., 2012; Galgano et al., 2008; Kedde et al., 2010; Miles et al., 2012). It was thus surprising to find that PUM1 directly regulates *ATXN1* mRNA stability without harnessing the miRNA regulatory system.

Equally unexpected was the discovery of a role for *Pum1* in the maintenance of proper brain structure and neurological function. We found *Pum1* expressed in all brain regions, but the deficits we observed in the heterozygous mice—progressive motor incoordination, hind-paw clasping, kyphosis, and Purkinje cell and dendritic degeneration—were reminiscent of the SCA1 mouse phenotype. The *Pum1* null mice phenocopied the SCA1 knockin mice but developed even more severe Purkinje cell pathology, showing significant neuronal loss after only 2 months. We believe this is explained at least in part by constitutive  $\sim 50\%$  increase in WT *Atxn1* in contrast with the gradual nature of the accumulation of polyglutamine-expanded proteins. Even in the SCA1 knockin mice,

the mutant protein takes time to accumulate to levels that produce symptoms. In the *Pum1* mutants, however, the levels of *Atxn1* are elevated from the very beginning of life.

The dramatic exacerbation of disease progression in SCA1 knockin mice lacking a copy of *Pum1* indicates a genetic interaction between *Pumilio1* and *Ataxin1*, but the precocity of disease symptoms in the double mutants could conceivably

highly conserved eight-nucleotide motif (Galgano et al., 2008). PUF proteins regulate mRNA stability by several mechanisms leading to mRNA instability or translational repression (Goldstrohm et al., 2006; Suh et al., 2009). One of the most well-studied mechanisms for PUM1 activity, however, involves the miRNA machinery: PUM1 modifies the secondary structure of the 3' UTR of its target mRNAs to allow regulation through specific

arise through either of two mechanisms: the loss of one copy of *Pum1* directly increasing levels of mutant *Atxn1*, or an additive effect ascribable to the combination of two severe mutations (loss of *Pum1* and the CAG expansion in *Atxn1*). Our results argue for the former possibility. First, the defects observed in *Pum1* mutant mice were largely corrected by reducing *Atxn1* levels. Second, the *Pum1*<sup>+/-</sup>;*Atxn1*<sup>+/-</sup> mice were healthier than either *Pum1*<sup>+/-</sup> or *Atxn1*<sup>+/-</sup> mice. Third, we found that viral overexpression of *Pum1* in the mouse brain reduced both the WT and expanded [154Q] forms of the *Atxn1* protein (Figure S6E). These three facts support the notion that the neurological deficits exhibited by the *Pum1* mutant mice are caused primarily by a rise in WT *Atxn1* levels, even though other *Pum1* targets are undoubtedly affected as well. Without knowing the full set of *Pum1* targets, we cannot rule out all other pathways, but we were able to evaluate levels of two well-studied PUM1 targets, E2F3 and p27, in *Pum1*<sup>+/-</sup> as well as *Pum1*<sup>+/-</sup>;*Atxn1*<sup>+/-</sup> mice. We did not find any significant change in the levels of these two proteins in the cerebella of either *Pum1*<sup>+/-</sup> or *Pum1*<sup>+/-</sup>;*Atxn1*<sup>+/-</sup> mice compared to other genotypes (data not shown).

It is well established that the severity of neurodegeneration in SCA1 correlates with the levels of expanded (or even WT) ATXN1 and that decreasing ATXN1 accumulation can reverse the disease phenotypes in SCA1 models (Fernandez-Funez et al., 2000; Park et al., 2013). It has been only relatively recently, however, that we have understood that in some neurodegenerative diseases, such as AD and PD, too much of the WT protein can produce the same phenotype as the mutant protein (Chartier-Harlin et al., 2004; Ibáñez et al., 2004; Rovelet-Lecrux et al., 2006; Rumble et al., 1989; Singleton et al., 2003). Our first study of transgenic mice overexpressing WT human ATXN1[30Q] under the control of a Purkinje-cell-specific promoter (*Pcp2-Atxn1*-(CAG)30Q) failed to reveal cerebellar pathology or ataxia (Burrigh et al., 1995), but our later work monitoring the mice throughout their lifespans revealed that *Pcp2-Atxn1*-(CAG)30Q mice develop mild Purkinje cell degeneration in later life (Fernandez-Funez et al., 2000). This suggested that only dramatic overexpression of WT ATXN1 would be neurotoxic, but the artificiality of the transgenic model—ATXN1 cDNA was massively and postnatally expressed only in cerebellar Purkinje cells and without the 3' UTR—limited its relevance for human patients. Here we introduce the evidence that a moderate (30%–60%) increase in the levels of endogenous WT *Atxn1*, expressed in the correct temporal and spatial pattern throughout the brain and preserving all its regulatory elements, is deleterious to neuronal function.

*Atxn1* is expressed throughout the brain (Figure S2A) from very early embryonic stages (Banfi et al., 1994, 1996; Servadio et al., 1995). The present study suggests that mutations in *PUM1* or copy-number changes in *ATXN1* could cause cerebellar neurodegeneration in humans by increasing the levels of ATXN1 throughout development. Variations in *PUM1* or other factors that govern ATXN1 levels could also underlie the individual differences in SCA1 onset for the same CAG repeat length.

In conclusion, we propose that identifying molecules capable of regulating ATXN1 levels provides insight into factors that contribute to cerebellar degeneration. We further propose that

studying factors that regulate the RNA stability of proteins such as APP, TAU, or  $\alpha$ -SYN might uncover candidate genes as well as binding sites whose mutation could lead to AD or PD—two diseases for which our understanding of molecular genetic causes is still very limited. For these and the ever-lengthening list of neurodegenerative conditions that do not fit Mendelian categories, it may prove most fruitful to search for factors that elevate the levels of key disease-driving proteins.

## EXPERIMENTAL PROCEDURES

### Bioinformatic Analysis

The ATXN1-3' UTR was downloaded from UTRdb (Grillo et al., 2010). The ATXN1 3' UTR was scanned against CoMeTa (Gennarino et al., 2012), HOCTARdb (Gennarino et al., 2011), and TargetScan (Friedman et al., 2009) to identify all putative miRNAs regulating ATXN1. The secondary structure of ATXN1-3' UTR was calculated with the Vienna RNAfold (Gruber et al., 2008) package by using default parameters in a Minimum Free energy "MFE" (Zuker and Stiegler, 1981) and Boltzmann ensemble "Centroid" (Hofacker and Stadler, 2006). ATXN1-3' UTR was scanned against all known RBP motifs downloaded from the database of RBP specificities (RBPDB) (Cook et al., 2011).

### Cell Culture and Transfection

The Human embryonic kidney immortalized 293 cells (HEK293T) were grown in DMEM (Invitrogen), supplemented with 10% of heat-inactivated fetal bovine serum (FBS) and penicillin/streptomycin. All cells were incubated at 37°C in a humidified chamber supplemented with 5% CO<sub>2</sub>. Transfection of HEK293T cells was performed using jetPRIME Transfection Reagent (Polyplus transfection) according to the manufacturer's protocol. Cells were transfected with 50 pmol of either miRIDIAN Dharmacon microRNA Mimics (miR-101a or negative control cel-miR-67) or Ambion small interfering RNA (siAGO2, siPUM1, or scramble-siRNA control). For overexpression studies, the full cDNA of *PUM1* (4,635 nt) was amplified by Platinum Taq DNA Polymerase High Fidelity (Invitrogen) and cloned into a mammalian expression vector termed pcDNA3.1(+) (Invitrogen). Cells were transfected with 0.5  $\mu$ g of either pcDNA3.1(+)-*PUM1* or control pcDNA3.1(+).

### RNA Extraction and Quantitative Real-Time PCR

HEK293T cells were seeded in 6-well plates before transfection. After 48 hr, cells were collected and processed for RNA extraction. Total RNA was obtained using the miRNeasy kit (Qiagen) according to the manufacturer's instructions. RNA extraction from mouse cerebrum or cerebellum was extracted from 75 mg of tissue. RNA was quantified using the NanoDrop 1000 (Thermo Fisher). Quality of RNA was assessed by gel electrophoresis. cDNA was synthesized using Quantitect Reverse Transcription kit (Qiagen) starting from 1  $\mu$ g of DNase-treated RNA. qRT-PCR experiments were performed using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories) with PerfeCta SYBR Green FastMix, ROX (Quanta Biosciences). Real-time PCR results were analyzed using the comparative Ct method normalized against the housekeeping gene *GAPDH* (Vandesompele et al., 2002). The range of expression levels was determined by calculating the standard deviation of the  $\Delta$ Ct (Pfaffl, 2001).

### Luciferase Assay

The full-length 3' UTR of human ATXN1 mRNA was subcloned into psiCHECK-2 vector (Promega) by XbaI and NheI restriction enzymes (Lee et al., 2008). The partial 3' UTR, containing binding sites 1 (582–782), 2 (2712–2912), and 3 (5175–5375), was amplified by PCR and cloned into psiCHECK-2 vector (Promega). Mutagenesis reactions were performed using the QuikChange XL Site-Directed Mutagenesis kit (Stratagene). Primers for mutagenesis analysis were automatically designed by QuikChange software (Stratagene). HEK293T cells in 24-well plates were transfected with 30 ng of psiCHECK-2 construct plus the following: 50 pmol of siPUM1 or control scramble-siRNA and 0.5  $\mu$ g of pcDNA3.1(+)-*PUM1* or control pcDNA3.1(+)



using Lipofectamine 2000 (Invitrogen). After 24 hr, luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

### Western Blot

HEK293T cells were seeded in 6-well plates before transfection. After 72 hr, cells were processed for protein extraction. For mouse tissues, the entire cerebrum and cerebellum were processed for protein extraction. Both pellet and tissues were lysed with RIPA buffer (25 mM Tris-HCl, pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS, and complete protease inhibitor cocktail [Roche]), then placed for 15 min on ice followed by centrifugation at 13,000 rpm at 4°C for 15 min. Proteins were quantified by Pierce BCA Protein Assay Kit (Thermo Scientific) and resolved by high-resolution Bolt 4%–12% Bis-Tris Plus Gel (Life Technologies) according to the manufacturer's instruction.

### RNA-CLIP

Brains from WT or *Pum1* knockout mice were dissected out, and cerebella were separated from the rest of the brains. After separation, the tissue was triturated in 8 ml of ice-cold HBSS until cells were evenly dissociated with no visible chunks. The cell suspension was layered on a chilled 10 cm sterile tissue culture plate and exposed to 150 mJ/cm<sup>2</sup> UVC (Stratagene, model UV Stratalinker 2400) on ice. After one exposure, the cell suspension was gently swirled and exposed again to UVC at 100 mJ/cm<sup>2</sup>. The cell suspension was pelleted for individual immunoprecipitation (IP). Cells were lysed in lysis buffer (50 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% sodium deoxycholate, 80 U/ml RNase OUT [Invitrogen] with protease inhibitor [Roche]). Soluble fractions were pre-cleared with protein A-sepharose beads, rabbit control IgG (Sigma), 0.05% BSA, and 0.2 µg/ml yeast tRNA (Invitrogen). Pre-cleared lysates were incubated with control IgG or *Pum1* (5 µg) (Bethyl Laboratories, see Antibodies) together with protein A-sepharose beads and incubated overnight at 4°C with gentle rotation. Next day, beads were washed five times with lysis buffer. Beads were treated using 20 units of RNase-free DNase (Roche) for 15 min at room temperature, followed by 50 µg proteinase K (Roche) treatment for 30 min at 37°C. Immunoprecipitated RNA was isolated using miRNeasy kit (Qiagen), and RT-PCR was performed using primers designed to amplify *Atxn1* cDNA regions upstream and downstream of the predicted *Pum1*-binding site (see Primers). Isolated RNA from a fraction (10%) of pre-cleared lysate was used as input.

### RNA Stability

Total RNA from HEK293T cells, RNA quality, cDNA synthesized, and qRT-PCR experiments were obtained as described above in [RNA Extraction and Quantitative Real-Time PCR](#). HEK293T cells were seeded in 24-well plates before transfection for both experiments (Figures 3E and 3F). For Figure 3E, the partial 3' UTR containing the *PUM1* WT or Mut binding sites 3 (5175–5375) were the same as used for the luciferase assay (see [Luciferase Assay](#) section). HEK293T cells were transfected with 30 ng of psiCHECK-2 vectors (Promega) containing *ATXN1*-3' UTR WT or Mut binding sites using Lipofectamine 2000 (Invitrogen). After 36 hr, cells were treated with 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) at the final concentration of 20 µg/ml, and total RNA for qRT-PCR analysis was collected at different time points. *Firefly* data were normalized to the respective *Renilla*. For Figure 3F, HEK293T cells were transfected with Ambion siRNA for *PUM1* (siPUM1) or Scramble (siScramble) at the final concentration of 40 nM. After 48 hr, cells were collected, and the total RNA was processed for qRT-PCR. All data were normalized to *GAPDH*. The housekeeping gene *GAPDH* was used to compare all the qRT-PCR values. All western blot experiments were performed as described in the [Western Blot](#) section.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.012>.

### AUTHOR CONTRIBUTIONS

V.A.G. and H.Y.Z. conceived the study, designed experiments, analyzed and interpreted data, and wrote the manuscript. V.A.G. performed molecular and behavioral tests. R.K.S., J.J.W., A.D.M., K.H., J.-Y.K., P.J.-N., A.D.R., and L.S.S. contributed to molecular work. H.K. performed bioinformatic analyses. T.A.C., H.T.O., and R.V.S. contributed to interpreting data.

### ACKNOWLEDGMENTS

We thank Dr. Haifan Lin (Yale Stem Cell Center and Department of Cell Biology, Yale University School of Medicine) for sharing the *Pum1* mutant mice; A. McCall for helpful suggestions on analysis; X. Liu for genotyping; C. Spencer for behavioral training; C. Ljungberg for ISH analysis; D. Yu for confocal microscopy training; and members of the H.Y.Z. laboratory for helpful discussions. We also thank A. Ballabio, T. Klisch, S. Yamamoto, M. Rousseaux, M. Sardiello, C. Schaaf, C. Alcott, and V. Brandt for helpful suggestions and critical reading of the manuscript. This research was supported by R01NS027699-26 (to H.Y.Z.) and the RNA In Situ Hybridization, Confocal and Mouse Behavioral Cores at the BCM Intellectual and Developmental Disabilities Research Center (IDDR; NIH/NICHHD P30-HD024064) and 1R01NS089664-01 (to R.V.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Received: July 22, 2014

Revised: December 1, 2014

Accepted: January 30, 2015

Published: March 12, 2015

### REFERENCES

- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805–810.
- Banfi, S., Servadio, A., Chung, M.Y., Kwiatkowski, T.J., Jr., McCall, A.E., Duvick, L.A., Shen, Y., Roth, E.J., Orr, H.T., and Zoghbi, H.Y. (1994). Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat. Genet.* 7, 513–520.
- Banfi, S., Servadio, A., Chung, M., Capozzoli, F., Duvick, L.A., Elde, R., Zoghbi, H.Y., and Orr, H.T. (1996). Cloning and developmental expression analysis of the murine homolog of the spinocerebellar ataxia type 1 gene (*Sca1*). *Hum. Mol. Genet.* 5, 33–40.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Bhattacharyya, S.N., Habermacher, R., Martine, U., Closs, E.I., and Filipowicz, W. (2006). Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125, 1111–1124.
- Bowman, A.B., Yoo, S.Y., Dantuma, N.P., and Zoghbi, H.Y. (2005). Neuronal dysfunction in a polyglutamine disease model occurs in the absence of ubiquitin-proteasome system impairment and inversely correlates with the degree of nuclear inclusion formation. *Hum. Mol. Genet.* 14, 679–691.
- Bowman, A.B., Lam, Y.C., Jafar-Nejad, P., Chen, H.K., Richman, R., Samaco, R.C., Fryer, J.D., Kahle, J.J., Orr, H.T., and Zoghbi, H.Y. (2007). Duplication of *Atxn1* suppresses SCA1 neuropathology by decreasing incorporation of polyglutamine-expanded ataxin-1 into native complexes. *Nat. Genet.* 39, 373–379.
- Burridge, E.N., Clark, H.B., Servadio, A., Matilla, T., Feddersen, R.M., Yunis, W.S., Duvick, L.A., Zoghbi, H.Y., and Orr, H.T. (1995). SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* 82, 937–948.
- Chartier-Harlin, M.C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Leveque, C., Larvor, L., Andrieux, J., Hulihan, M., et al. (2004). Alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* 364, 1167–1169.

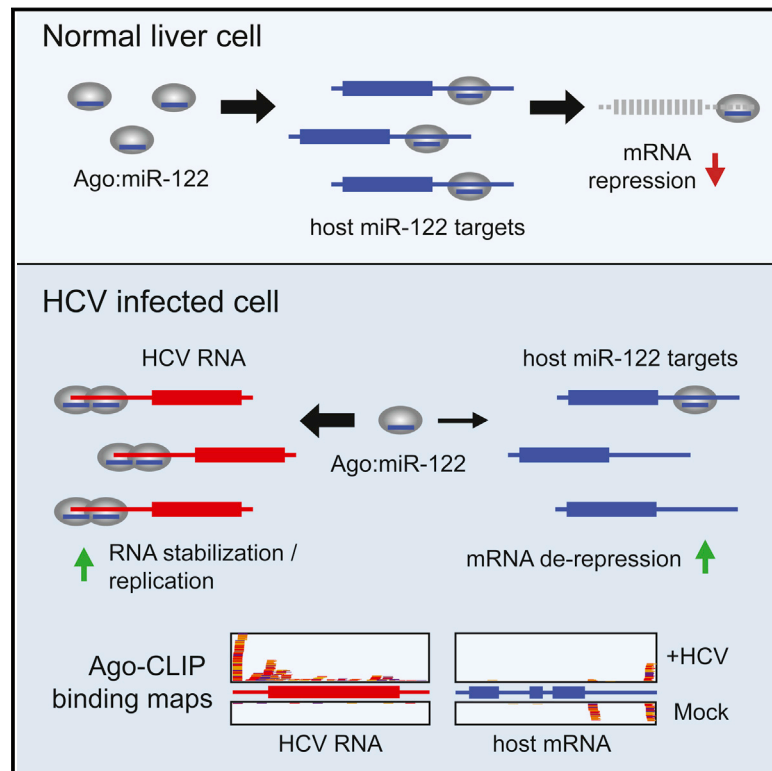


- Chen, D., Zheng, W., Lin, A., Uyhazi, K., Zhao, H., and Lin, H. (2012). Pumilio 1 suppresses multiple activators of p53 to safeguard spermatogenesis. *Curr. Biol.* 22, 420–425.
- Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.* 39 (Database issue), D301–D308.
- Cummings, C.J., Reinstein, E., Sun, Y., Antalffy, B., Jiang, Y., Ciechanover, A., Orr, H.T., Beaudet, A.L., and Zoghbi, H.Y. (1999). Mutation of the E6-AP ubiquitin ligase reduces nuclear inclusion frequency while accelerating polyglutamine-induced pathology in SCA1 mice. *Neuron* 24, 879–892.
- Fabian, M.R., and Sonenberg, N. (2012). The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.* 19, 586–593.
- Fernandez-Funez, P., Nino-Rosales, M.L., de Gouyon, B., She, W.C., Luchak, J.M., Martinez, P., Turiegano, E., Benito, J., Capovilla, M., Skinner, P.J., et al. (2000). Identification of genes that modify ataxin-1-induced neurodegeneration. *Nature* 408, 101–106.
- Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.
- Friend, K., Campbell, Z.T., Cooke, A., Kroll-Conner, P., Wickens, M.P., and Kimble, J. (2012). A conserved PUF-Ago-eEF1A complex attenuates translation elongation. *Nat. Struct. Mol. Biol.* 19, 176–183.
- Galgano, A., Forrer, M., Jaskiewicz, L., Kanitz, A., Zavolan, M., and Gerber, A.P. (2008). Comparative analysis of mRNA targets for human PUF-family proteins suggests extensive interaction with the miRNA regulatory system. *PLoS ONE* 3, e3164.
- Gennarino, V.A., Sardiello, M., Mutarelli, M., Dharmalingam, G., Maselli, V., Lago, G., and Banfi, S. (2011). HOCTAR database: a unique resource for microRNA target prediction. *Gene* 480, 51–58.
- Gennarino, V.A., D'Angelo, G., Dharmalingam, G., Fernandez, S., Russolillo, G., Sanges, R., Mutarelli, M., Belcastro, V., Ballabio, A., Verde, P., et al. (2012). Identification of microRNA-regulated gene networks by expression analysis of target genes. *Genome Res.* 22, 1163–1172.
- Glorian, V., Mailliot, G., Polès, S., Iacovoni, J.S., Favre, G., and Vagner, S. (2011). HuR-dependent loading of miRNA RISC to the mRNA encoding the Ras-related small GTPase RhoB controls its translation during UV-induced apoptosis. *Cell Death Differ.* 18, 1692–1701.
- Goldstrohm, A.C., Hook, B.A., Seay, D.J., and Wickens, M. (2006). PUF proteins bind Pop2p to regulate messenger RNAs. *Nat. Struct. Mol. Biol.* 13, 533–539.
- Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V.A., Horner, D.S., Pavesi, G., Picardi, E., and Pesole, G. (2010). UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* 38 (Database issue), D75–D80.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* 36 (Web Server issue), W70–W74.
- Haass, C., and Selkoe, D.J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat. Rev. Mol. Cell Biol.* 8, 101–112.
- Hofacker, I.L., and Stadler, P.F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics* 22, 1172–1176.
- Ibáñez, P., Bonnet, A.M., Débarges, B., Lohmann, E., Tison, F., Pollak, P., Agid, Y., Dürr, A., and Brice, A. (2004). Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. *Lancet* 364, 1169–1171.
- Kedde, M., van Kouwenhove, M., Zwart, W., Oude Vrielink, J.A., Elkon, R., and Agami, R. (2010). A Pumilio-induced RNA structure switch in p27-3' UTR controls miR-221 and miR-222 accessibility. *Nat. Cell Biol.* 12, 1014–1020.
- Kim, H.H., Kuwano, Y., Srikantan, S., Lee, E.K., Martindale, J.L., and Gorospe, M. (2009). HuR recruits let-7/RISC to repress c-Myc expression. *Genes Dev.* 23, 1743–1748.
- Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y., and Orr, H.T. (1998). Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell* 95, 41–53.
- Kundu, P., Fabian, M.R., Sonenberg, N., Bhattacharyya, S.N., and Filipowicz, W. (2012). HuR protein attenuates miRNA-mediated repression by promoting miRISC dissociation from the target RNA. *Nucleic Acids Res.* 40, 5088–5100.
- Lee, Y., Samaco, R.C., Gatchel, J.R., Thaller, C., Orr, H.T., and Zoghbi, H.Y. (2008). miR-19, miR-101 and miR-130 co-regulate ATXN1 levels to potentially modulate SCA1 pathogenesis. *Nat. Neurosci.* 11, 1137–1139.
- Leeb, M., Dietmann, S., Paramor, M., Niwa, H., and Smith, A. (2014). Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* 14, 385–393.
- Lukong, K.E., Chang, K.W., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. *Trends Genet.* 24, 416–425.
- Martin, I., Kim, J.W., Lee, B.D., Kang, H.C., Xu, J.C., Jia, H., Stankowski, J., Kim, M.S., Zhong, J., Kumar, M., et al. (2014). Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. *Cell* 157, 472–485.
- Miles, W.O., Tschöp, K., Herr, A., Ji, J.Y., and Dyson, N.J. (2012). Pumilio facilitates miRNA regulation of the E2F3 oncogene. *Genes Dev.* 26, 356–368.
- Orr, H.T., Chung, M.Y., Banfi, S., Kwiakowski, T.J., Jr., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P., and Zoghbi, H.Y. (1993). Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.* 4, 221–226.
- Park, J., Al-Ramahi, I., Tan, Q., Mollema, N., Diaz-Garcia, J.R., Gallego-Flores, T., Lu, H.C., Lagalwar, S., Duvick, L., Kang, H., et al. (2013). RAS-MAPK-MSK1 pathway modulates ataxin 1 protein levels and toxicity in SCA1. *Nature* 498, 325–331.
- Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29, e45.
- Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. *Nat. Med. Suppl.* 10, S10–S17.
- Rovelet-Lecrux, A., Hannequin, D., Raux, G., Le Meur, N., Laquerrière, A., Vital, A., Dumanchin, C., Feuillette, S., Brice, A., Vercelletto, M., et al. (2006). APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.* 38, 24–26.
- Rumble, B., Retallack, R., Hilbich, C., Simms, G., Multhaup, G., Martins, R., Hockey, A., Montgomery, P., Beyreuther, K., and Masters, C.L. (1989). Amyloid A4 protein and its precursor in Down's syndrome and Alzheimer's disease. *N. Engl. J. Med.* 320, 1446–1452.
- Servadio, A., Koshy, B., Armstrong, D., Antalffy, B., Orr, H.T., and Zoghbi, H.Y. (1995). Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nat. Genet.* 10, 94–98.
- Singleton, A.B., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). alpha-Synuclein locus triplication causes Parkinson's disease. *Science* 302, 841.
- Soto, C. (2003). Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* 4, 49–60.
- Spassov, D.S., and Jurecic, R. (2003). The PUF family of RNA-binding proteins: does evolutionarily conserved structure equal conserved function? *IUBMB Life* 55, 359–366.
- Suh, N., Crittenden, S.L., Goldstrohm, A., Hook, B., Thompson, B., Wickens, M., and Kimble, J. (2009). FBF and its dual control of *gld-1* expression in the *Caenorhabditis elegans* germline. *Genetics* 181, 1249–1260.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, H0034.
- Wan, Y., Qu, K., Zhang, Q.C., Flynn, R.A., Manor, O., Ouyang, Z., Zhang, J., Spitale, R.C., Snyder, M.P., Segal, E., and Chang, H.Y. (2014). Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709.

- Wang, X., McLachlan, J., Zamore, P.D., and Hall, T.M. (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110, 501–512.
- Watase, K., Weeber, E.J., Xu, B., Antalffy, B., Yuva-Paylor, L., Hashimoto, K., Kano, M., Atkinson, R., Sun, Y., Armstrong, D.L., et al. (2002). A long CAG repeat in the mouse Sca1 locus replicates SCA1 features and reveals the impact of protein solubility on selective neurodegeneration. *Neuron* 34, 905–919.
- Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. *Trends Genet.* 18, 150–157.
- Zoghbi, H.Y., and Orr, H.T. (2000). Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* 23, 217–247.
- Zoghbi, H.Y., and Orr, H.T. (2009). Pathogenic mechanisms of a polyglutamine-mediated neurodegenerative disease, spinocerebellar ataxia type 1. *J. Biol. Chem.* 284, 7425–7429.
- Zuker, M., and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* 9, 133–148.

# Hepatitis C Virus RNA Functionally Sequesters miR-122

## Graphical Abstract



## Authors

Joseph M. Luna, Troels K.H. Scheel, ..., Charles M. Rice, Robert B. Darnell

## Correspondence

ricec@rockefeller.edu (C.M.R.),  
darnelr@rockefeller.edu (R.B.D.)

## In Brief

Hepatitis C virus uniquely requires the liver-specific tumor suppressor miRNA, miR-122, for its replication. During infection, viral RNA specifically sequesters miR-122 to de-repress its normal host targets, which may facilitate the long-term oncogenic potential of HCV.

## Highlights

- Genome-wide miRNA binding profiles were elucidated for HCV infection
- HCV RNA functionally reduces miR-122 binding on endogenous mRNA targets
- HCV miRNA sponging can be redirected by swapping viral miRNA tropism
- Modeling validates single-cell measurements of HCV-induced mRNA de-repression

## Accession Numbers

GSE64680



# Hepatitis C Virus RNA Functionally Sequesters miR-122

Joseph M. Luna,<sup>1,2</sup> Troels K.H. Scheel,<sup>1,3</sup> Tal Danino,<sup>1,4</sup> Katharina S. Shaw,<sup>1</sup> Aldo Mele,<sup>2</sup> John J. Fak,<sup>2</sup> Eiko Nishiuchi,<sup>1</sup> Constantin N. Takacs,<sup>1,5</sup> Maria Teresa Catanese,<sup>1,8</sup> Ype P. de Jong,<sup>1,6</sup> Ira M. Jacobson,<sup>6</sup> Charles M. Rice,<sup>1,\*</sup> and Robert B. Darnell<sup>2,7,\*</sup>

<sup>1</sup>Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, NY 10065, USA

<sup>2</sup>Laboratory of Molecular Neuro-Oncology and Howard Hughes Medical Institute, The Rockefeller University, New York, NY 10065, USA

<sup>3</sup>Copenhagen Hepatitis C Program (CO-HEP), Department of Infectious Disease and Clinical Research Centre, Copenhagen University Hospital, Hvidovre, and Department of International Health, Immunology and Microbiology, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2200 Copenhagen N, Denmark

<sup>4</sup>Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>5</sup>Laboratory of Cellular Biophysics, The Rockefeller University, New York, NY 10065, USA

<sup>6</sup>Center for the Study of Hepatitis C, Division of Gastroenterology and Hepatology, Weill Cornell Medical College, New York, NY 10065, USA

<sup>7</sup>New York Genome Center, 101 Avenue of the Americas, New York, NY 10013, USA

<sup>8</sup>Present address: Department of Infectious Diseases, King's College London School of Medicine, Guy's Hospital, London Bridge, London SE1 9RT, UK

\*Correspondence: [ricec@rockefeller.edu](mailto:ricec@rockefeller.edu) (C.M.R.), [darnellr@rockefeller.edu](mailto:darnellr@rockefeller.edu) (R.B.D.)

<http://dx.doi.org/10.1016/j.cell.2015.02.025>

## SUMMARY

Hepatitis C virus (HCV) uniquely requires the liver-specific microRNA-122 for replication, yet global effects on endogenous miRNA targets during infection are unexplored. Here, high-throughput sequencing and crosslinking immunoprecipitation (HITS-CLIP) experiments of human Argonaute (AGO) during HCV infection showed robust AGO binding on the HCV 5'UTR at known and predicted miR-122 sites. On the human transcriptome, we observed reduced AGO binding and functional mRNA de-repression of miR-122 targets during virus infection. This miR-122 “sponge” effect was relieved and redirected to miR-15 targets by swapping the miRNA tropism of the virus. Single-cell expression data from reporters containing miR-122 sites showed significant de-repression during HCV infection depending on expression level and site number. We describe a quantitative mathematical model of HCV-induced miR-122 sequestration and propose that such miR-122 inhibition by HCV RNA may result in global de-repression of host miR-122 targets, providing an environment fertile for the long-term oncogenic potential of HCV.

## INTRODUCTION

Hepatitis C virus (HCV) is a hepatotropic positive-strand RNA virus of the Flaviviridae family that is a leading cause of liver disease globally, with morbidities such as fibrosis, cirrhosis, and hepatocellular carcinoma (Yamane et al., 2013). The long ORF of the ~9.6 kb HCV genome encodes a polyprotein processed

into ten proteins and is flanked by critical structured UTRs. Unique to this virus is a dependence on the liver-specific microRNA-122 (miR-122) (Jopling et al., 2005). Whereas miRNAs typically interact with the 3'UTRs of mRNAs to promote mRNA destabilization and/or translational repression (Bartel, 2009), the binding of miR-122 to two binding sites (seed site S1 and S2) in the 5'UTR of HCV genomic RNA is critical for viral replication (Jopling et al., 2008; Machlin et al., 2011) by moderately stimulating viral protein translation (Henke et al., 2008) and, in concert with Argonaute (AGO), by stabilizing and protecting the uncapped HCV RNA genome from degradation (Li et al., 2013b; Sedano and Sarnow, 2014; Shimakami et al., 2012). As the predominant miRNA in the liver, miR-122 has multiple roles to regulate lipid metabolism (Esau et al., 2006), iron homeostasis (Castoldi et al., 2011), and circadian rhythms (Gatfield et al., 2009). MiR-122 knockout studies in vivo have revealed potent anti-inflammatory and anti-tumorigenic functions (Hsu et al., 2012; Tsai et al., 2012). Antagonizing miR-122 as an HCV therapeutic is a novel strategy (Lanford et al., 2010) with the first-in-class inhibitor, miravirsin/SPC3649, currently in phase II clinical studies (Janssen et al., 2013).

Studies of miRNA action during virus infections have been enhanced with the advent of high-throughput methods to elucidate genome-wide miRNA:mRNA interaction networks biochemically. Such methods (Chi et al., 2009; Hafner et al., 2010), broadly relying on cross-linking and immunoprecipitation (CLIP) of RNA bound to protein, have been applied to latent Kaposi's sarcoma-associated herpesvirus (KSHV) (Haecker et al., 2012) and Epstein Barr virus (EBV) infections to uncover miRNA regulatory networks involved in promoting viral latency (Skalsky et al., 2012) and regulating cellular apoptosis (Riley et al., 2012).

In the current study, we elucidated global miRNA:target interaction maps during HCV infection on host and viral RNA. We observed AGO engagement at the HCV 5'UTR miR-122 sites, describe replication-dependent argonaute binding throughout



viral genomic RNA, and provide evidence of miR-122 binding on an HCV resistant to miR-122 antagonism. On the host transcriptome, our results revealed globally reduced AGO binding and specific de-repression of miR-122 targets upon virus infection. This surprising systems-level observation suggests that HCV RNA functionally sequesters miR-122, and exhibits a miRNA “sponge” effect analogous to roles proposed for competing endogenous RNAs (ceRNA) (Salmena et al., 2011). Taken together, our results establish an RNA virus as a specific and indirect regulator of miRNA activity in the cell.

## RESULTS

### Argonaute HiTS-CLIP of HCV Infected Cells

To study miRNA interactions during HCV infection, we either electroporated RNA or infected Huh-7.5 hepatoma cells with J6/JFH1-Clone2 HCV and after 48–72 hr, when most cells were infected, performed AGO-CLIP and RNA-seq measurements (Figures S1A–S1C). AGO-CLIP was performed using linker ligation as previously described (Figures S1D–S1F) (Moore et al., 2014). Alignment statistics for CLIP datasets presented in this paper are summarized in Tables S2, S3, S4, and S5.

Due to known linker ligation biases in the preparation of small RNA libraries (Zhuang et al., 2012), we used polyG tailing (adapted from Ingolia et al., 2009) to determine miRNA abundance profiles (Figure S1G), and found that miR-122 at ~4.9% is the seventh most abundant miRNA (Figure S1H and Table S1). This correlated with previous data on miR-122 abundance in these cells (Figures S1I and S1J). No systematic bias from linker ligation was observed on mRNA targets due to the relative heterogeneity of RNaseA cleavage in creating mRNA AGO footprints (Figures S1K and S1L). For subsequent analysis on mRNA-CLIP clusters, we focused on searching the top 50 seed families derived from poly-G CLIP studies, which constituted over 97% of miRNAs identified in Huh-7.5 cells.

### An AGO Binding Map of HCV RNA Confirms Extensive miR-122 Engagement

To define a small RNA interaction map on HCV and human mRNA, CLIP reads were mapped onto the HCV and human genomes. Among the 1%–2% of CLIP reads mapping to HCV, AGO binding sites were identified by clustering overlapping reads and identifying statistically significant peaks above a uniformly distributed background (Darnell et al., 2011; Licatalosi et al., 2012). We observed major peaks in the 5'UTR, E1, E2, NS5A, and NS5B regions of the genome (Figure 1A, top). No significant binding was observed on the negative strand (data not shown). Notably, 50% of all AGO binding events on HCV RNA overlapped the known miR-122 seed sites in the 5'UTR (Figures 1A and 1B). Achieving nucleotide precision on the AGO:mRNA crosslink site via crosslink-induced mutation site analysis (CIMS [Zhang and Darnell, 2011]), we observed an enrichment of crosslink sites predominantly within and immediately upstream of S2 (at positions 28 and 35) and, to a lesser degree, S1 base pairing locations (Figure 1B). The second largest peak was observed in the HCV IRES and overlapped the pseudoknot and coding start site (Figure 1D). No canonical 7-mer or 8-mer binding sites for the top 50 miRNA seeds were noted in this region; however, a

putative non-canonical miR-122 site in the IRES (Pang et al., 2012) may explain the observed AGO binding.

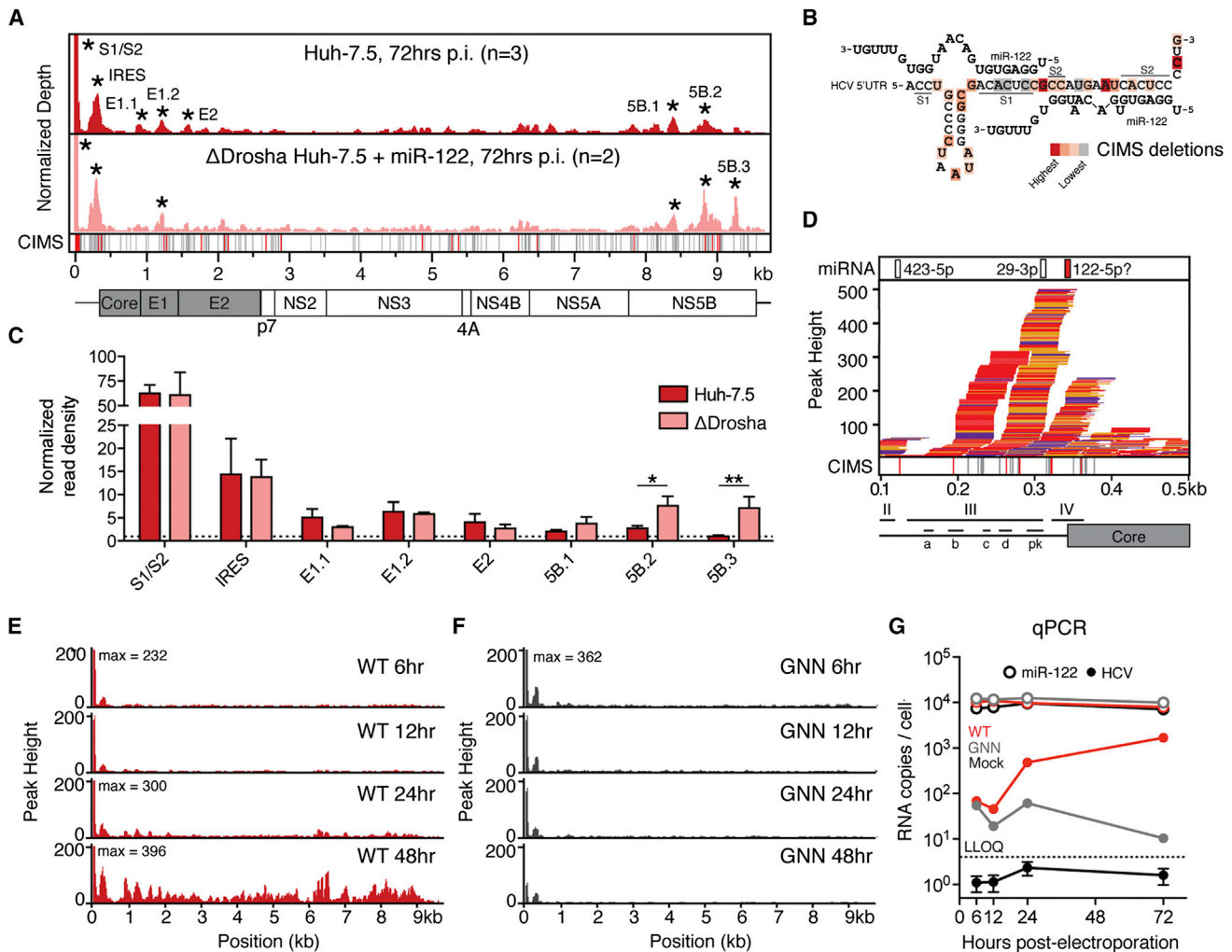
To probe the general miRNA dependence of AGO binding on HCV RNA and to specifically enrich for miR-122-dependent binding, we deleted Drosha from Huh-7.5 cells to globally disrupt most miRNA biogenesis, using a CRISPR-based genome editing strategy (Figure S2). From CLIP in  $\Delta$ Drosha supplemented with miR-122 mimic to support HCV infection (Figure S2G), we observed that binding to S1/S2 and the IRES was maintained, and notably enhanced at NS5B binding sites (Figures 1A and 1C), both of which contain conserved miR-122 sites (Figure S3), of which one was previously shown to be inhibitory for HCV replication (Nasheri et al., 2011). AGO binding to E1 and E2 peaks was reduced in  $\Delta$ Drosha cells, suggesting that a minor proportion of AGO binding on HCV RNA is due to other miRNAs. While reports have suggested that numerous miRNAs interact with HCV RNA (reviewed in Singaravelu et al., 2014), among these only let-7 and miR-196 families fell within the top 50 miRNAs expressed, and no seeds from either of these families were observed within significant AGO binding peaks. Taken together, these data suggest that miR-122 constitutes the predominant miRNA interaction with HCV RNA in these cells and is largely confined to the 5'UTR.

### AGO Binding to miR-122 Sites on HCV RNA Occurs Early and Is Replication Independent

To address the timing of AGO binding to HCV RNA, we performed CLIP over a time course after electroporation of WT or replication defective (GNN) RNA genomes. We observed comparable AGO binding at the S1/S2 sites of WT and GNN mutants as early as 6 hr post-electroporation (Figures 1E and 1F). AGO binding to the 5'UTR in general, and to miR-122 sites in particular, remained stable throughout the WT time course but decreased steadily for the GNN mutant. AGO binding to regions outside the 5'UTR emerged after 24 hr, were not observed in the GNN mutant, and correlated with HCV RNA abundance over time (Figures 1E–1G). This suggested early AGO binding to the 5'UTR and additional replication or abundance-dependent low-level AGO targeting of the viral ORF.

### An HCV Resistant to miR-122 Antagonism Engages miR-122 and AGO

To further dissect the impact of non-miR-122 AGO binding on HCV RNA, we focused on an HCV recombinant that is resistant to miR-122 antagonism (Li et al., 2011). This virus, for which the first HCV 5'UTR stem-loop is replaced with cellular U3 snoRNA, lacks the S1 site but contains an intact S2 site. Upon deleting miR-122 from Huh-7.5 cells ( $\Delta$ miR-122) using a CRISPR-based strategy (Figure S2), we observed that while WT virus replication was abolished in  $\Delta$ miR-122 cells, U3 virus replication was largely unaffected (Figures 2A and 2B). Reintroducing miR-122 completely rescued WT virus replication and had a small but consistent (2- to 4-fold) proviral effect on the U3 virus (Figures 2A and 2B), suggesting that U3 virus replication is largely miR-122 independent. U3 virus replication could also be launched in  $\Delta$ Drosha cells, yet in WT cells S2 p3 and p3,4 mutants were negative for HCV replication over 3 weeks, suggesting perturbation of overlapping functions on the RNA (data not shown). In CLIP on U3



**Figure 1. Argonaute Binding Maps on HCV RNA**

(A) Mock-subtracted binding map of AGO-CLIP reads across HCV genomic RNA in WT or  $\Delta$ Drosha Huh-7.5 cells. Data were normalized to total cellular and virus read depth for comparison. Significant peaks per track are named by location and indicated by asterisks. Bottom CIMS track shows location of all deletions (gray) and statistically significant CIMS deletions (red) from the WT track.

(B) AGO binding in significant peaks from WT Huh-7.5 cells in (A) shown as normalized read densities calculated per dataset. Data were normalized to background read density of non-peak regions (dashed line). Asterisks,  $^{*}p < 0.01$ ,  $^{**}p < 0.05$ , Student's *t* test. Error bars,  $\pm$  SD.

(C) Schematic of a miR-122 binding model to S1 and S2 highlighting locations of CIMS deletions.

(D) Zoom in view of AGO binding from WT cells in (A) across the viral IRES into the coding sequence. IRES domains (II-IV), associated stemloops (a-d), and the pseudoknot (pk) region are indicated. Upper track displays seeds for the top 50 miRNA seeds, previously proposed miR-122 binding (Pang et al., 2012) highlighted in red.

(E and F) AGO binding time course of WT (E) and replication deficient (GNN) (F) HCV post-electroporation (*n* = 2).

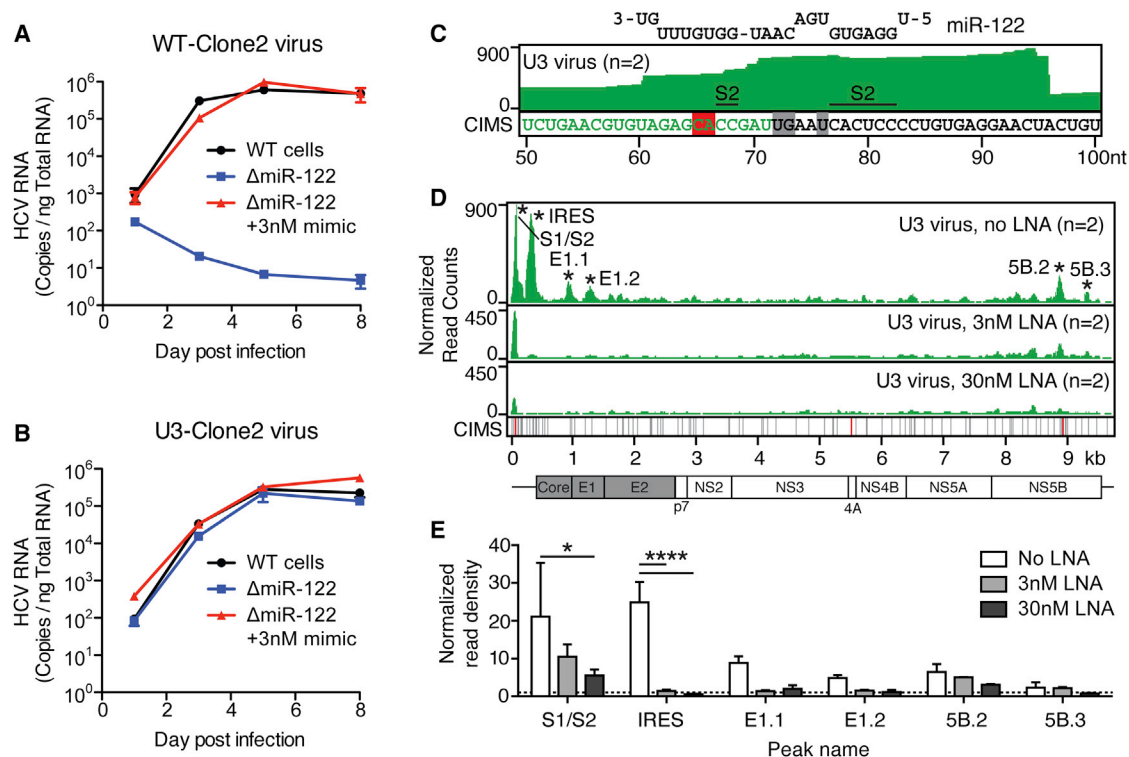
(G) Absolute qPCR measurements of miR-122 and HCV RNA levels at indicated time points post-electroporation (*n* = 3). Replication-deficient J6/JFH1-GNN and mock controls are shown. Dashed line indicates lower limit of quantitation. Error bars,  $\pm$  SD. See also Figure S3.

infected Huh-7.5 cells, we observed AGO binding and crosslink mapping at the S2 miR-122 site specifically (Figure 2C) mirroring WT virus and indicating that the U3 viral RNA residually engages miR-122. In the presence of increasing concentrations of miR-122 locked nucleic acid (LNA) inhibitor, we observed a dose-dependent decrease in AGO binding across the viral ORF, and a significant decrease in the S2 and IRES binding locations (Figures 2D and 2E), consistent with a limited proviral effect of miR-122 and specific miR-122 binding to S2 and the IRES. Furthermore, S2 and IRES binding was lost in U3 infected

$\Delta$ Drosha and  $\Delta$ miR-122 cells (data not shown). The striking miR-122 independence of this virus points to a potential avenue of resistance to LNA-based therapeutics in the form of recombinant viruses with similarly large 5'UTR stem-loops.

### HCV Infection Functionally Reduces Argonaute Binding on Host miR-122 Targets

Given the crucial requirement of miR-122 for HCV replication, and in light of the result that HCV RNA levels accumulate to within one log of miR-122 levels (Figure 1G), we hypothesized that the HCV



**Figure 2. An HCV Mutant Resistant to miR-122 Antagonism Engages AGO and miR-122**

(A and B) Time course qPCR measurements of WT-Clone2 virus (A) or U3-Clone2 virus (B) in WT cells or in ΔmiR-122 cells with or without 3 nM miR-122 supplementation. Error bars, ± SD.

(C) AGO binding map (top track) and CIMS locations (bottom track) across the U3 virus 5'UTR corresponding to miR-122 binding at S2. Relevant CIMS deletions are shown in gray (not significant) and red (significant). U3 snoRNA sequence is shown in green.

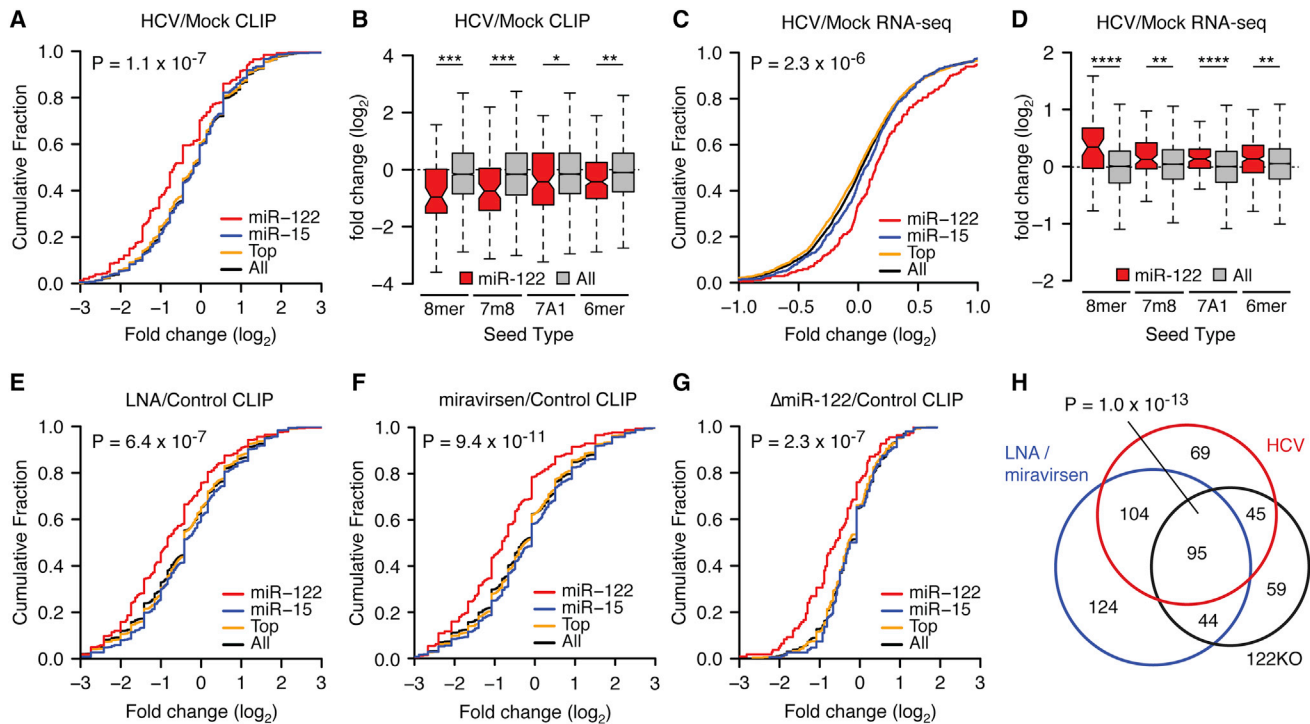
(D) AGO binding map across the U3 virus genome after treatment with increasing doses of LNA122. Significant peaks are named by location and indicated by asterisks. Bottom CIMS track shows location of all deletions (gray) and significant CIMS deletions (red) for the untreated dataset.

(E) AGO binding in significant peaks from untreated U3 datasets in (D) shown as normalized read densities calculated per dataset. \*\*\*\*p < 0.0001, \*p < 0.05, one-way ANOVA with bonferroni correction. Error bars, ± SD.

genome may act as a “sponge” for cellular miR-122, where viral replication may exert a broadly de-repressive effect on host miR-122 targets. We reasoned that this effect would be noticeable via CLIP as reduced AGO binding of miR-122 targets upon infection that may result in a specific increase in the expression of these targets measured by mRNA-seq. Indeed, comparing HCV infected to uninfected cells, we observed significantly reduced AGO binding globally for mRNA targets for which a miR-122 seed was present, compared to the combined targets of the miR-15/16 family, as a representative targetome of similar size to the miR-122 target network, and the top 10 or the top 50 miRNA families cumulatively (Figure 3A). Significant changes in miR-122 binding were observed for all canonical seed types (as defined in Bartel, 2009) (Figure 3B). Additionally, the greatest change in AGO association was observed in CLIP clusters within 3'UTRs and was less significant in CDS (coding exons), 5'UTRs, and introns (Figure S4A). Notably, the 3'UTR targets of other miRNAs suggested to bind HCV RNA directly were not altered upon HCV infection (Figure S4B). Through RNA-seq measurements we observed functional de-repression of CLIP-derived miR-122 3'UTR targets after virus infection such that greater RNA abundance was evident when compared to all miRNA tar-

gets (Figure 3C). Likewise, we observed significant expression changes for all miR-122 target seed types (Figure 3D). Compared to bioinformatic prediction using Targetscan6.2 (TS) (Lewis et al., 2005), we found that CLIP largely complemented and expanded upon predicted miR-122 targets (Figures S4C–S4G). 3'UTR targets identified via CLIP and predicted by TS exhibited the greatest change in AGO binding (Figure S4C) and mRNA de-repression (Figures S4D and S4E) compared to expressed targets unique to either search modality. Of the expressed 731 miR-122 CLIP targets of all seed types identified via CLIP, 48% and 9% overlapped with non-conserved and conserved TS predictions, respectively (Figure S4F). Focusing on a more stringent set of 7-mer and 8-mer seeds for CLIP data yielded even greater overlap, such that only 5% of CLIP-derived targets were not represented in either TS conservation category (Figure S4G). These results highlight a broad convergence between CLIP and bioinformatic prediction to outline a set of miR-122 targets specifically de-repressed upon virus infection.

To further corroborate our CLIP observations with HCV infection, we performed CLIP after pharmacologic inhibition of miR-122 and in ΔmiR-122 Huh-7.5 cells. The reduced AGO binding on miR-122 3'UTR targets during HCV infection was similar to



**Figure 3. HCV Infection De-Represses Endogenous miR-122 Targets**

(A) Cumulative density function (CDF) of the log<sub>2</sub> fold change in CLIP binding between infected and uninfected cells for all 3'UTR clusters containing indicated 7- to 8-mer seeds by family, from triplicate experiments. "Top" refers to the top 10 miRNA families, exclusive of miR-122. "All" refers to the top 50 miRNA families, inclusive of miR-122. Two-sided K-S test p value between miR-122 and all targets shown.

(B) The mean log<sub>2</sub> fold change (± ranges) in CLIP binding on miR-122 3'UTR targets versus all targets during HCV infection broken down by seed type.

(C) A CDF plot during HCV infection as in (A) but measuring target mRNA expression via RNA-Seq, from duplicate experiments at 72 hr post-infection. Targets with more than one miRNA binding site were collapsed such that no gene is represented more than once per category.

(D) The mean log<sub>2</sub> fold change (± ranges) in mRNA expression of CLIP targets during HCV infection broken down by seed type.

(E–G) CDF plot as in (A), between treatment over control cells with LNA122 (E) or miravirsin (F) at 30 nM or genetic deletion (G) of miR-122 (ΔmiR-122), each from triplicate experiments.

(H) Proportional Venn diagram showing the overlap of miR-122 targets with reduced CLIP binding across ΔmiR-122, LNA or miravirsin treatment, and HCV infection conditions. Hypergeometric p value of overlap shown.

Asterisks: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05, two-sided Mann-Whitney U-test. See also Figures S2 and S4.

30 nM LNA122 or miravirsin treatment (Figures 3E and 3F) and to ΔmiR-122 cells compared to unedited controls (Figure 3G). Interrogating the list of 3'UTR targets exhibiting reduced AGO binding across these three conditions revealed highly significant overlap (Figure 3H), suggesting that the effect of HCV replication on lowering functional miR-122 levels is functionally similar to antagonizing miR-122. The full complement of miRNA targets identified in these studies is presented in Table S6.

### Transcriptome Regulation by miR-122 Sequestration In Vitro Is Predictive of Sequestration In Vivo

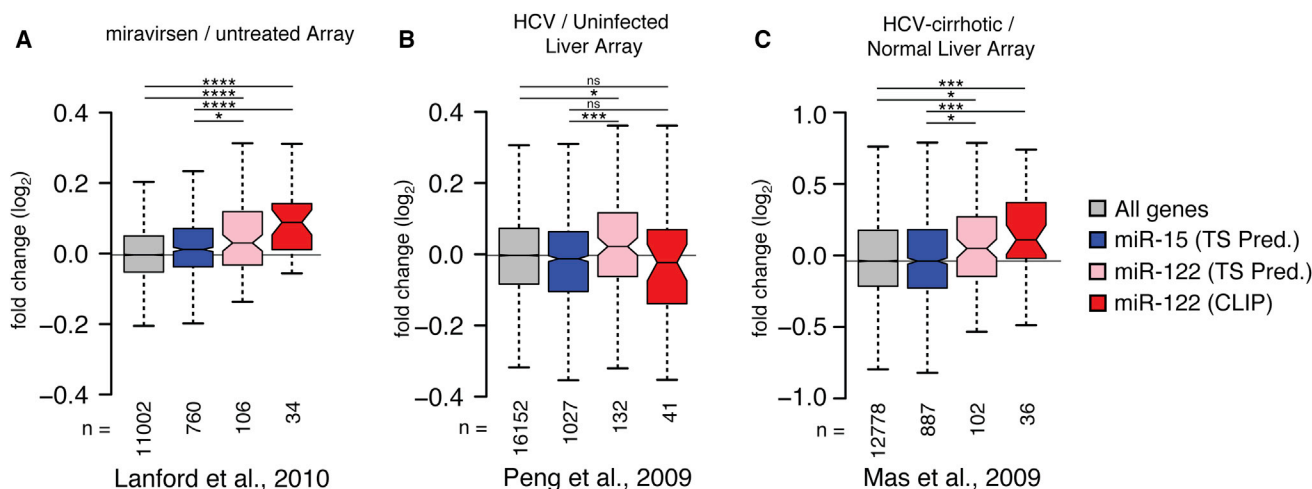
As the global effect of HCV replication on host miR-122 usage mirrored LNA inhibition, and more profoundly miR-122 deletion, we hypothesized that miR-122 targets as a class might be de-repressed in human livers as a result of HCV infection. To this end, we performed a meta-analysis of published liver biopsy microarray data related to miR-122 inhibition or HCV infection. Comparing TS predictions for expressed miR-122 or miR-15 targets, as well as CLIP-identified miR-122 targets to all expressed genes in microarray data from miravirsin-treated chimpanzee

livers (Lanford et al., 2010), we noted a significant de-repression for TS or CLIP miR-122 targets compared to all genes or to predicted miR-15 targets (Figure 4A). In this dataset, the CLIP-identified miR-122 targetome as a group was broadly more de-repressed than TS predictions. We performed the same analysis comparing HCV infected versus uninfected samples from two array datasets (Mas et al., 2009; Peng et al., 2009) and, despite the unknown proportion of infected cells, found in both that miR-122 target predictions were significantly de-repressed compared to all genes or to miR-15 target predictions in both datasets (Figures 4B and 4C). While these results cannot directly confirm an HCV sponge effect in vivo, they do emphasize that the overlap between CLIP results in vitro and expression results in vivo may indicate specific de-repression of miR-122 targets during HCV infection.

### Validation of HCV-Induced miR-122 Sequestration in Bulk and Single Cells

The results thus far describe the global characteristics of the HCV-induced miR-122 sponge effect on the host transcriptome.





**Figure 4. Meta-Analysis of Published Array Data Suggests HCV-Induced Changes on the miR-122 Target Network**

(A) Miravirsin pre- and post-treatment array data from four HCV infected chimpanzees (Lanford et al., 2010) was binned according to conserved 7- to 8-mer TargetScan (TS) predictions for miR-15 or miR-122, or from miR-122 targets with CLIP support from the current study. Boxplot whiskers denote 1.5 times the inter-quartile distance from the nearest quartile. The mean fold change in expression for miR-122 targets was compared to miR-15 targets or all genes represented on the array, where the number of genes in each bin (n) is indicated.

(B) Analysis as in (A) comparing 24 HCV-positive to 5 -negative liver biopsies (Peng et al., 2009).

(C) Analysis as in (A) comparing 41 HCV-positive with cirrhosis samples to 19 normal livers (Mas et al., 2009).

Asterisks: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05, ns p > 0.05, two-sided Mann-Whitney U-test.

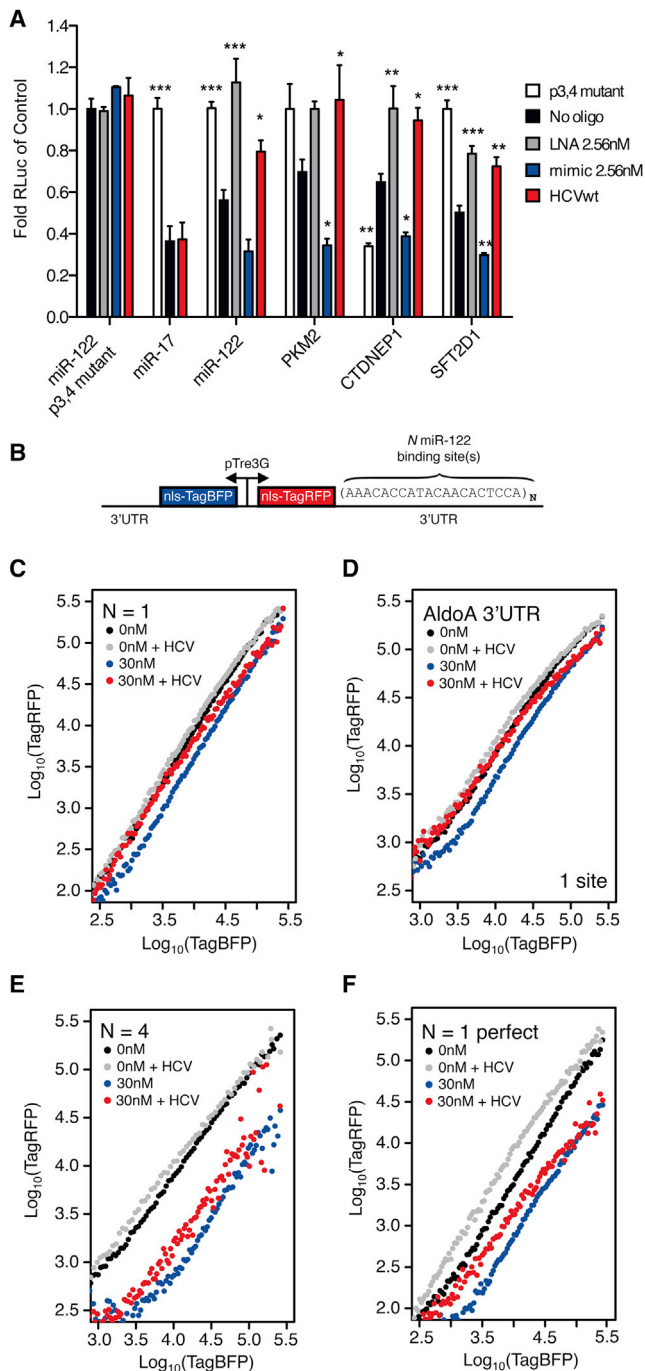
We next used luciferase and fluorescent reporters of miRNA activity to validate the HCV miR-122 sponge on individual 3'UTRs. The endogenous repression of luciferase reporters containing one miR-122 seed, as well as 3'UTRs of CLIP-derived miR-122 targets was enhanced upon adding additional miR-122 and was reversed upon LNA-mediated miR-122 inhibition (Figure 5A). We also observed seed-dependent statistically significant de-repression of these reporters upon virus infection (Figure 5A).

As cellular mRNA and HCV RNA expression levels can vary widely between individual cells (Kandathil et al., 2013; Sheahan et al., 2014), we sought to achieve a more thorough understanding of the HCV miRNA sponge on host miRNA targets at a quantitative single-cell level. Previous work demonstrated that miRNAs generate thresholds of gene expression such that miRNA repression can be highest on low abundance targets and can be virtually non-existent on high-abundance targets. Furthermore, these thresholds can be altered upon manipulating miRNA levels (Mukherji et al., 2011). To test whether HCV replication could broadly impact functional miR-122 levels, we adapted the strategy used by Mukherji et al. to construct two-color tet-inducible fluorescent reporters of miRNA activity amenable to flow cytometry (Mukherji et al., 2011) (Figure 5B).

Testing reporters with N = 1 and 6 miR-122 binding sites in the presence of miR-122 mimic, we observed miR-122 repression that increased with N, as expected, whereas adding LNA122 decreased repression (Figure S5A–S5D). HCV infection in both contexts resembled LNA inhibition where de-repression was notably more pronounced in cells expressing low amounts of reporter, demonstrating previously reported miRNA thresholding effects (Mukherji et al., 2011) (Figures S5A and S5C). Importantly,

no such changes were observed for a reporter with a p3,4 miR-122 seed (“N1m”). (Figures S5E and S5F). Additionally, we tested a reporter with a perfectly complementary miR-122 site, thus making the miRNA behave as an siRNA. This reporter exhibited no thresholding such that mimic repression, or LNA and HCV de-repression was observed at all expression levels (Figures S5G and S5H). These data suggest that HCV infection modulates functional miR-122 levels to relieve endogenous repression on host targets in a stoichiometric manner, governed by target expression level and the number of miRNA binding sites.

As miR-122 levels in Huh7-derived cells are estimated to be 10-fold lower than primary adult liver tissue (Chang et al., 2004), we next explored the HCV sponge effect in the presence of excess miR-122. Exogenous miR-122 addition increased intracellular miR-122 by up to 10-fold in Huh-7.5 cells, within the range of miR-122 levels measured from patient liver biopsies (Figures S6A and S6B). As no changes in HCV RNA levels were observed, the resulting miR-122:HCV ratio went from ~15:1-fold at the lowest, to over 100:1 with 30 nM of miR-122 mimic added (Figure S6C). Testing N = 1 or AldoA 3'UTR reporter constructs in this in vivo-like context, we observed that HCV infection was able to relieve 30 nM of mimic repression to untreated levels for low but not high abundance targets (Figures 5C and 5D). The ability for HCV to rescue excess miR-122 repression was not as pronounced for the N = 4 construct (Figure 5E) whereas a reporter containing a perfectly complementary miR-122 site was particularly sensitive to rescue by HCV replication (Figure 5F). Similar, dose-dependent results were obtained under 0.3 or 3 nM mimic treatment for all constructs (Figures S6D–S6G). Taken together, these results suggest that miR-122 sponging by HCV can exist in more physiologic miR-122 concentration settings.



**Figure 5. Validation of HCV-Induced De-Repression of miR-122 Targets in Bulk and Single-Cell Resolution**

(A) Luciferase reporter measurements for synthetic miR-122, miR-17, or cellular 3'UTR target constructs. Data were normalized to "no oligo" p3,4 mutant conditions. Significance testing was performed relative to endogenous "no oligo" repression for each tested construct. Error bars,  $\pm$  SEM. Asterisks: \*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05, ANOVA with Bonferroni correction.

(B) Two-color fluorescent reporter containing a bidirectional Tet promoter that drives expression of blue and red fluorescent proteins (TagBFP and TagRFP). Each fluorescent protein is tagged with a nuclear localization sequence (NLS) to aid in flow cytometric analysis. The 3'UTR of TagRFP is

### A Quantitative Model of miR-122 Sponging by HCV RNA

To achieve a more quantitative understanding of the HCV sponge, we used our dose-dependent mimic and LNA reporter system measurements to expand the miRNA model of gene regulation presented by Mukherji et al. to incorporate a competing self-replicating viral target (Figure 6A). Here, if HCV RNA is present at sufficiently high numbers or has relatively high-binding strengths compared to other miR-122 targets, it acts to reduce the available miR-122 pool, and de-represses miR-122 targets ( $r$ , measured as TagRFP fluorescence) relative to non-targets ( $r_0$ , measured as TagBFP fluorescence) (Figure 6B). We developed a quantitative formula for HCV-induced reduction of the miR-122 pool in this scenario (see Supplemental Information). Assuming steady-state levels of HCV RNA at the time of measurements resulted in a decrease of the model parameter  $\theta$ , which governs the amount of free miRNA in the system. The number of miR-122 sites is estimated by the model parameter  $\lambda$  that is related to the total binding strength of miR-122 to a particular site. By tuning these parameters, we accurately fitted experimental data of endogenous miR-122 repression of reporters with increasing numbers of miR-122 sites (Figure 6C).

To explore the effect of HCV on the miR-122 pool, we fitted the model to experimental data with four miR-122 sites during infection (Figure 6D), and estimated the change to parameter  $\theta$  to correspond to an approximate 50% reduction in available miR-122. A similar result was obtained for the N = 4 construct in the presence of HCV and 30 nM miR-122 mimic (Figure 6E). The model estimated that the highest theoretical HCV levels reducing the miR-122 pool by 90% could de-repress mRNA targets by up to 4.5-fold for low-expressed mRNAs (Figure 6F). Synthetic reporter measurements agreed with model predictions for 50% reductions in miR-122 levels, where de-repression was most drastic for low expressed targets harboring multiple miR-122 sites (Figure 6G).

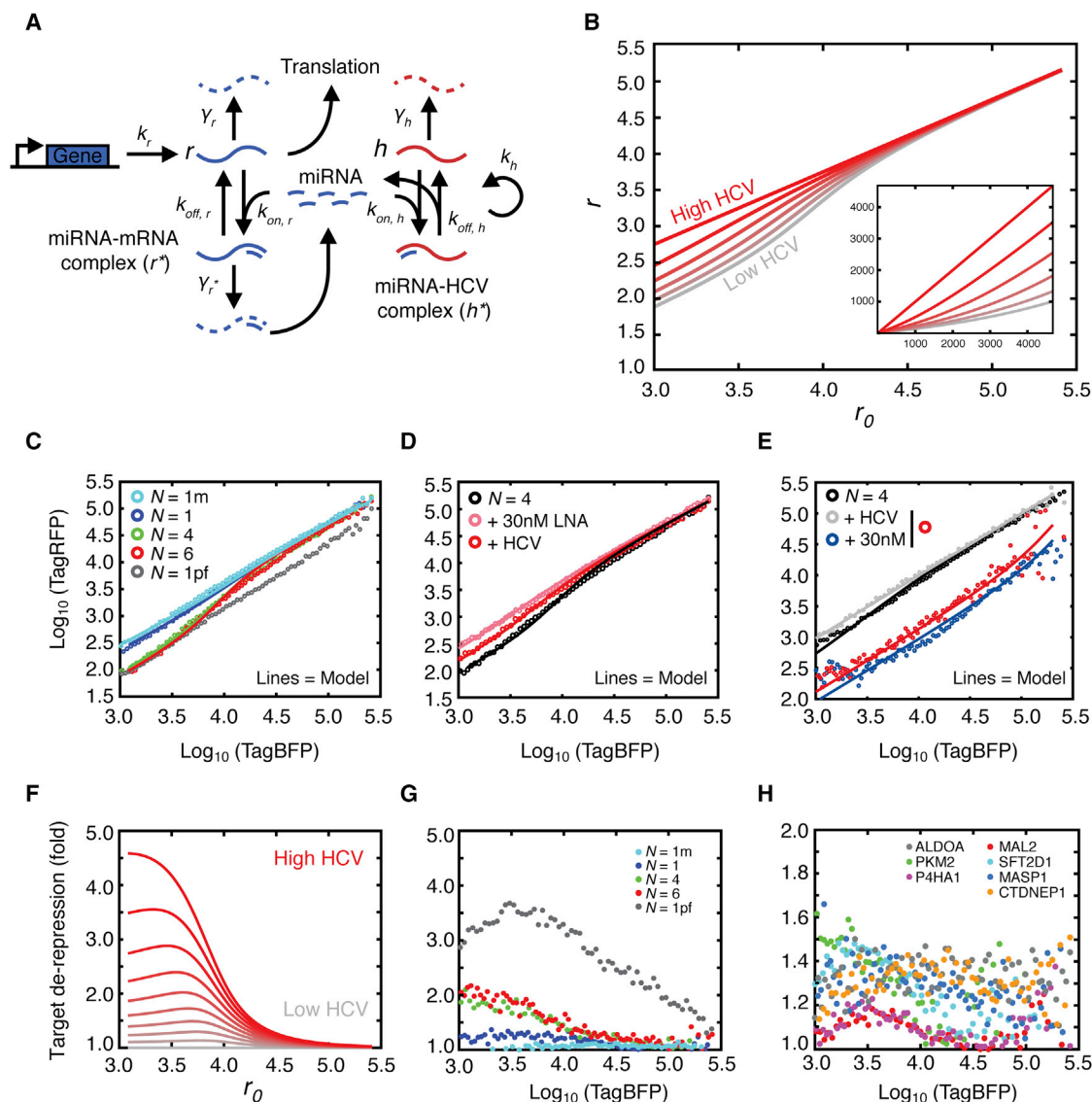
In similar measurements with reporters containing full 3'UTRs, we observed modest de-repression upon HCV infection with an average change of 25% across all expression levels for the previously known targets with one miR-122 site, AldoA, PKM2, and P4HA1, but not for CS 3'UTRs (Figures 6H and S5H–S5L). The novel CLIP-identified targets CTDNEP1, SFT2D1, MASP1 and MAL2 behaved similarly, with all four tested being reduced upon miR-122 mimic addition and all except MAL2 de-repressed upon adding virus (Figures S5M–S5P). Our quantitative model outlines several factors controlling HCV-induced de-repression of host mRNA targets, namely, the expression level of the target mRNA, mRNA-miR-122 binding strength, and the number of sites on the target mRNA.

### A miR-15-Dependent HCV Redirects miRNA Sequestration

To go beyond the correlative connection between our CLIP, RNA-seq, and modeling results, we set out to confirm the

engineered to contain  $N$  binding sites for miR-122, or full 3'UTRs of miR-122 targets.

(C–F) Log-log transfer functions for N = 1 (C), ALDOA 3'UTR (D), N = 4 (E) or one perfectly complementary (F) miR-122 site in the presence or absence of 30 nM miRNA mimic and/or HCV infection.



**Figure 6. Quantitative Modeling of miR-122 Sequestration by HCV**

(A) Illustration of model reactions for miR-122 dynamics, including transcription and translation of a target mRNA, binding to miR-122 and decay of mRNA species. HCV RNA can replicate, be degraded, or bind miR-122, and functionally sequestering miR-122 and leading to de-repression of mRNA targets.

(B) Increasing amounts of HCV or a relative increase in binding strength at miR-122 sites leads to changes in single-cell gene expression as compared to unregulated targets, with stronger effects at the low mRNA expression levels. Parameters used are fitted from data in (C). Each curve, from bottom to top represents a 20% reduction in the available miRNA pool by HCV. Inset displays model on a linear scale.

(C) Model fitting of the steady-state approximation to experimental data while increasing the number of binding sites corresponding to changes in total binding strength.

(D) Model fitting for the  $N = 4$  case showing a 50% reduction in the miRNA pool by HCV modeled by a proportional change in the theta parameter.

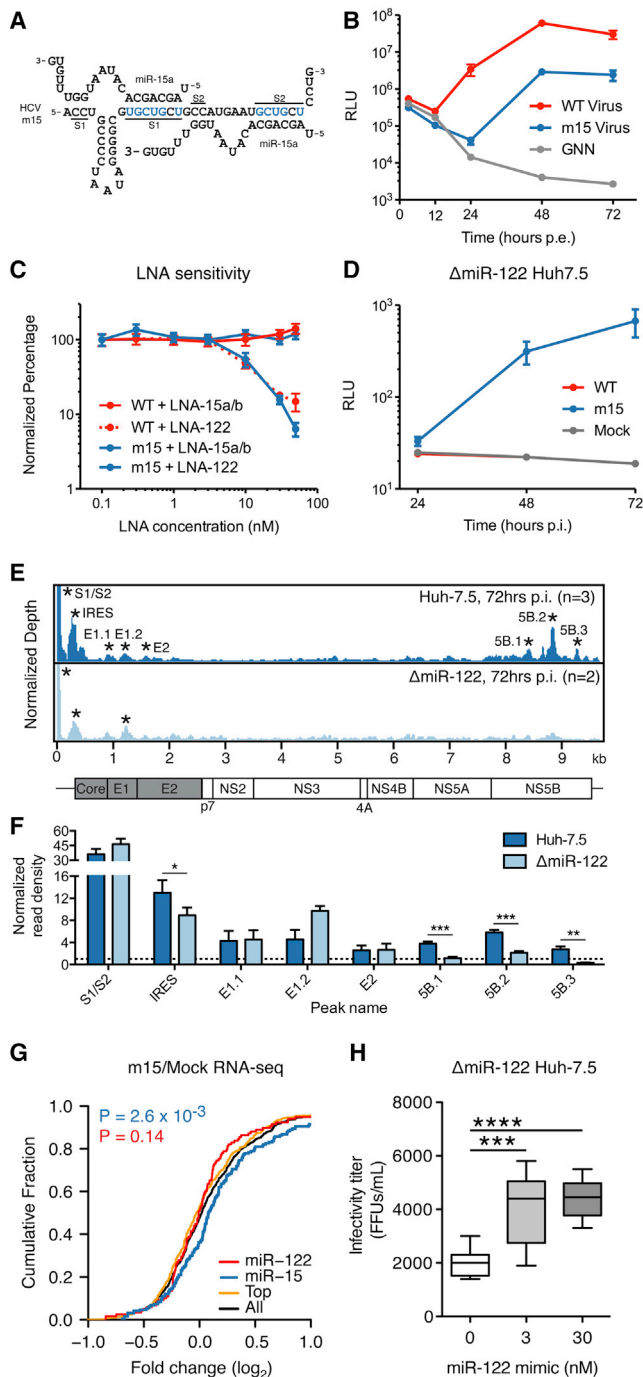
(E) Model fitting for the  $N = 4$  construct under 30 nM miR-122 mimic addition  $\pm$  HCV infection.

(F) Increasing HCV:miR-122 binding strength or HCV RNA abundance in the model results in functional de-repression of miR-122 targets. Each curve, from bottom to top, represents a 10% decrease in the available miR-122 pool by HCV.

(G and H) Experimental HCV induced derepression of synthetic miR-122 binding site constructs (G) or endogenous 3' UTRs with miR-122 binding sites (H). See also Figure S5 and S6.

sponge effect by swapping the miRNA tropism of the virus to determine if the miR-122 sponge could be redirected to the targets of another miRNA. We selected miR-15a/b, as these miRNAs had a sufficiently altered but GC-rich seed, and maintained the auxiliary pairing at nucleotides 2–3 and 30–31 of the

viral genome (Machlin et al., 2011) (Figures 7A and S7A). In total, miR-15a/b constituted  $5.3\% \pm 1.2\%$  of miRNA identified via miRNA-CLIP, compared to  $4.9\% \pm 2.0\%$  for miR-122 (Table S1 and Figure S7A). An electroporated miR-15 variant HCV luciferase reporter virus (m15) was viable, and replicated to within



**Figure 7. Exchanging HCV miRNA Tropism Redirects Functional miRNA Sequestration**

(A) Base pairing diagram of miR-15a onto the mutated m15 HCV RNA. Base changes from the WT at S1 and S2 are highlighted in blue.

(B) Luciferase measurements of supernatants from WT and m15 HCV reporter virus electroporations ( $\pm$  SD). Non-replicating GNN control is shown.

(C) Dose response of WT and m15 reporter viruses following pre-treatment with LNA inhibitors of miR-122 or miR-15a/b at indicated concentrations, measured at 96 hr post-infection ( $\pm$ SD).

(D) Time course post-infection of  $\Delta$ miR-122 Huh-7.5 cells of indicated viruses ( $\pm$ SD).

one log of the WT virus after 72 hr (Figure 7B). Notably, the miR-15 virus was resistant to increasing concentrations of LNA122 ( $IC_{50} > 50$  nM) but susceptible to LNA15a/b ( $IC_{50} = 10$  nM) (Figure 7C). Similar results were obtained with non-reporter viruses through measuring HCV replication, spread of infection and virus titers (Figures S7B–S7G). Unlike WT virus, the m15 virus was viable in  $\Delta$ miR-122 cells (Figure 7D), thus demonstrating complete HCV replication independent of miR-122.

AGO binding on the m15 virus in Huh-7.5 cells largely mirrored results observed with WT virus (Figure 7E, top). Interestingly, we observed reduced AGO binding in  $\Delta$ miR-122 cells for the m15 virus in IRES and NS5B peaks harboring conserved miR-122 seeds, further suggesting miR-122 dependence for AGO binding in these regions (Figure 7E and 7F).

Turning our attention to the host, we measured the effect on the miR-15 targetome due to m15 virus replication via RNA-seq and confirmed de-repression specific to miR-15 targets while no longer observing effects on the miR-122 target network (Figure 7G). These results highlight the causal nature of an HCV-induced miRNA sponge as both functional and somewhat modular. We note that the m15 virus sponge effect was generally weaker than for the WT virus, likely due to the lower replication level observed and possibly to binding of the miR-15 family member, miR-16, which shares the seed site but may not be able to engage the m15 genome due to lack of auxiliary pairing (Figure S7A). Modification of the HCV 5'UTR could have direct effects on viral replication, and may explain the slight attenuation observed for the m15 virus (Figure 7B).

To further evaluate pro- or anti-viral effects of miR-122 abundance on the m15 virus, we measured m15 virus replication in  $\Delta$ miR-122 cells after re-introducing miR-122 at various concentrations. We observed a slight but significant increase in secreted m15 virus levels upon re-introducing miR-122 (Figure 7H). RNA levels and percent of infected cells were also increased in m15 infected  $\Delta$ miR-122 cells upon restoring miR-122 levels (Figures S7H and S7I). Conversely, we observed a significant reduction in viral titers after LNA-122 treatment of m15 infected Huh-7.5 cells (Figure S7G). Taken together, these results suggest that miR-122 sequestration by HCV replication confers a slightly less pro-viral cellular environment; an effect

(E) AGO binding map of m15 virus infection in WT Huh-7.5 (top) or  $\Delta$ miR-122 Huh-7.5 cells (bottom). Data were normalized to total cellular read depth for cross track comparison. Statistically significant peaks per track are named by location and are indicated by asterisks.

(F) AGO binding in significant peaks from (E) shown as normalized read densities calculated per dataset. Two-sided Student's t test used. Error bars,  $\pm$  SD.

(G) CDF plot of the  $\log_2$  fold change in mRNA expression between HCV m15 infected and uninfected cells for all 3'UTR clusters containing indicated 7- to 8-mer seeds by family, from duplicate experiments at 96 hr post-infection. "Top" refers to the top 10 miRNA families, exclusive of miR-122 and miR-15. "All" refers to the top 50 miRNA families, inclusive of miR-122 and miR-15. Two-sided K-S test p value comparing miR-15 (blue) or miR-122 (red) clusters to "All" is shown.

(H) Infectivity titers of m15 virus in  $\Delta$ miR-122 Huh-7.5 cells complemented with exogenous miR-122 at indicated concentrations. One-way ANOVA with Bonferroni correction, whiskers,  $\pm$  ranges.

Asterisks: \*\*\*\*p < 0.0001, \*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05. See also Figure S7.



that may be a small trade-off for the hugely beneficial role miR-122 plays directly in the viral life-cycle.

## DISCUSSION

The aim of this study was to obtain an unbiased view of the miRNA interactome during HCV infection on both viral and cellular targets. We provide direct biochemical evidence of replication independent AGO association with HCV RNA at the 5'UTR in a number of infection contexts. Our results chiefly establish HCV RNA may act as a competitive inhibitor of miR-122 activity, an idea closely aligned with proposed roles for ceRNAs (Salmena et al., 2011). While both HCV RNA and ceRNAs share the theme of de-repressing a miRNA regulated network by increasing the pool of available targets through RNA expression, they differ in their mode of interaction with miRNAs. HCV genomic RNA critically requires miR-122 interaction to stabilize the viral genome and stimulate translation and replication, while most cellular miRNA targets are degraded upon encountering a miRNA. Moreover, unlike cellular mRNA targets, HCV genomic RNA is its own substrate for replication, and thus constitutes a positive feedback loop to sequester additional miR-122. These distinguishing features suggest different parameters for HCV versus ceRNA-based sponge effects on a miRNA target network. This is relevant in light of recent findings showing that endogenous miR-122 repression is only relieved when ceRNAs are forcibly expressed at super-physiological levels (Denzler et al., 2014). As HCV RNA represented at most between 1%–2% of CLIP reads following infection, our data suggest ceRNA activity may occur naturally at lower expression levels. Indeed, single-cell reporter measurements and mathematical modeling of the mRNA, HCV, and miRNA interplay suggest that HCV is able to de-repress host miR-122 targets, due to an approximate 2-fold reduction in available miR-122 under our experimental levels of HCV replication. Natural viral-derived miRNA sponges have been described previously (Cazalla et al., 2010; Lee et al., 2013) though no examples are currently known from RNA viruses. Whether this systems-level phenomenon occurs with other, more robust RNA virus infections remains to be explored.

The establishment of a miR-15-dependent HCV suggests that the miR-122 sponge effect is largely dispensable for the virus in the Huh-7.5 cell context. Indeed, we observed that LNA-122 slightly reduced m15 virus titers and that restoring miR-122 in  $\Delta$ miR-122 cells increased titers, suggesting that the miR-122 sponge may reflect a trade-off for the large, positive, and direct impact of miR-122 on WT HCV replication. Conceivably, HCV replication may exert enough pressure on miR-122 levels to de-repress targets such that the cellular environment is passively altered to negatively impact viral replication. Or more actively, there may exist miR-122 targets that act as sensors for low miR-122 levels, and by extension, the health of the hepatocyte. While future work will be needed to shed light on specific players involved in this process, our data suggest that viral replication faces a ceiling by reducing levels of an otherwise pro-viral miRNA.

How might the HCV miR-122 sponge impact a hepatocyte? Work with miR-122 knockout mice, which develop progressive liver disease that spontaneously results in HCC (Hsu et al.,

2012; Tsai et al., 2012) suggests that miR-122 tumor suppressor activity is essential for long-term liver homeostasis. It is tantalizing to speculate that miR-122 sequestration in a chronic HCV infection may be a molecular link to the heterogeneous liver dysfunction that characterizes HCV-induced disease. Indeed, a number of miR-122 targets that we confirm or establish via CLIP and reporter measurements, such as P4HA1, PKM2, and MASP1 are known to be upregulated in fibrosis or HCC (Jung et al., 2011; Li et al., 2013a), with MASP1 notable for being specifically linked with HCV-associated HCC (Saeed et al., 2013). Still, there are challenges for a direct demonstration of an HCV miR-122 sponge in vivo. Our bulk cell measurements estimate 10-fold higher miR-122 levels in primary liver tissue versus hepatoma cell line derivatives ( $10^5$  versus  $10^4$  estimate ranges per cell), while HCV levels per cell are estimated to range from 1 to  $10^2$  copies per hepatocyte (Kandathil et al., 2013), in contrast to Huh7 derivatives with  $10^3$  copies per cell. Yet, we observed functional HCV-induced de-repression in our single-cell reporter assay after addition of miR-122 in Huh-7.5 cells to mimic levels in vivo. Furthermore, we found specific de-repression of miR-122 targets in microarrays from HCV infected livers and from miravirsin-treated chimpanzees, supporting the existence of an HCV sponge effect in vivo. Unlike the >90% infection frequencies in Huh-7.5 cells, the percentage of infected hepatocytes in chronically infected patients based on in situ hybridization of liver biopsy tissue, ranges from as low as 0.07% to as high as 100%, with medians in the 20%–40% range (Liang et al., 2009; Pal et al., 2006). Combined with HCV genotype, dynamic replication variation within the liver, and host variability in innate immune responses (Sheahan et al., 2014), a complex picture of HCV infection emerges that would largely mask observations of HCV sponge effects in bulk cell or tissue AGO-CLIP measurements. As it remains possible that functional effects of such a sponge may impact highly infected cells, our data highlight the possibility of searching for transcriptome level changes to the miR-122 target network in response to HCV infection in individual cells. The extension of CLIP and RNA-seq in single-cell and primary contexts provides a compelling platform to address these and other long-term disease driven changes to a miRNA target network.

## EXPERIMENTAL PROCEDURES

### Culture of Cell Lines, Generation and Characterization of WT and m15 HCV

Cell lines were cultured and generated as described in Supplemental Information. pJ6/JFH1-Clone2, pJ6/JFH-Clone2-5AB-Ypet, and pJc1FLAG(p7-nsGluc2A) are fully infectious HCV non-reporter and reporter viruses, respectively, that have been previously described (Catanese et al., 2013; Horwitz et al., 2013; Marukian et al., 2008). To construct miR-15-dependent viruses in both backgrounds, we used an overlap PCR mutagenesis strategy. HCV cloning, RNA transcription, electroporation, infection, and related virus assays are described in detail in the Supplemental Information.

### Argonaute HITS-CLIP Analysis

Argonaute CLIP was performed generally following previous work (Chi et al., 2009), (Moore et al., 2014). Poly-G CLIP is a direct adaptation of the single linker ligation BrdU CLIP protocol (Weyn-Vanhenryck et al., 2014). Relevant details pertaining to the CLIP protocol, multiplexed library preparation, and bioinformatic analysis are described in full in the Supplemental Information.

### mRNA-Seq Library Construction

mRNA-seq libraries were prepared from Trizol extracted RNA following Illumina TruSeq protocols for poly-A selection, fragmentation, and adaptor ligation. Multiplexed libraries were sequenced as 100 nt single-end runs on either HiSeq-2000 or MiSeq platforms.

### Luciferase Reporter Assays

Luciferase reporter vectors were cloned by inserting short oligonucleotides or PCR amplified target 3'UTRs into psiCHECK-2. The sequence of DNA oligonucleotides used for cloning are found in Table S7. Huh-7.5 cells were transfected over night with 2.56 nM final concentration LNA122 (Exiqon) or miR-122 mimic (Thermo Fisher) using RNAi/MAX (Invitrogen). Alternatively, cells were infected with HCV (J6/JFH1-clone2), MOI = 3 overnight. Twenty-four hours later, cells were transfected with 1 ng/well psiCHECK-2 reporter plasmid using Lipofectamine2000 (Invitrogen) and incubated over night before lysis in Passive Lysis Buffer and evaluation of luciferase levels using the Dual Luciferase Reporter Assay (Promega) on a Omega Fluorostar reader (BMG Labtech).

### Single-Cell Reporter Measurements

Construction of miR-122 fluorescent reporters largely mirrored previous work with miR-20 (Mukherji et al., 2011). For a list of primers used in plasmid construction, please refer to Table S7. For flow cytometry, cells were run on a MACSQuant VYB flow cytometer (Miltenyi Biotec) after fixation to detect TagBFP, TagRFP, and Ypet signals. The raw FACS data were analyzed with FlowJo software to gate single, intact cells according to their forward (FSC-A) and side (SSC-A) scatter profiles. HCV-positive cells were gated on the basis of Ypet signal above uninfected background. Untransfected cells were used to characterize the cellular autofluorescence in BFP and RFP channels, from which we subtracted the mean plus two SD of the autofluorescent signal for each channel in transfected cells. Cells with BFP and RFP fluorescence levels less than 0 after background subtraction were excluded from further analyses. Data were log-transformed and binned according to BFP levels, and the mean RFP signal was calculated for each BFP bin.

### ACCESSION NUMBERS

HiTS-CLIP and RNA-seq data have been deposited in the GEO under accession number GSE64680.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.025>.

### AUTHOR CONTRIBUTIONS

J.M.L. conceived the project; J.M.L., T.K.H.S., C.M.R., and R.B.D. designed the experiments; J.M.L., T.K.H.S., K.S.S., A.M., J.J.F., E.N., and M.T.C. performed the experiments; C.N.T. constructed Huh-7.5 TetON cells; Y.P.J. and I.M.J. provided clinical samples; J.M.L. and T.D. developed the mathematical model; J.M.L., T.K.H.S., and T.D. analyzed the data; J.M.L., T.K.H.S., T.D., C.M.R., and R.B.D. wrote the manuscript with input from all authors.

### ACKNOWLEDGMENTS

We thank C. Zhang for bioinformatics support; M. Moore for flow cytometry assistance; the Rockefeller Genomics Resource Center; and M.E. Castillo, S.M. Pecoraro Di Vittorio, J. Palarca, J. Sable, B. Hough-Loomis, and J. Smith for excellent technical and administrative assistance. Lastly, we thank members of the Darnell and Rice labs for invaluable advice. This study was supported by grants from the Public Health Service, NIH, NIAID (AI099284, AI072613, AI075099, AI091707, AI090055), NINDS (NS034389, NS081706), Office of the Director through the NIH Roadmap for Medical Research (DK085713), NCI (CA057973), The Rockefeller University Center for Clinical

and Translational Science (UL1RR024143), the Center for Basic and Translational Research on Disorders of the Digestive System through the generosity of the Leona M. and Harry B. Helmsley Charitable Trust, the Greenberg Medical Research Institute, and the Starr Foundation. J.M.L. was supported by a David Rockefeller Graduate Student Fellowship. T.K.H.S. was supported by a Postdoctoral Fellowship and a Sapere Aude Research Talent Award from The Danish Council for Independent Research. C.N.T. was supported by an HHMI International Predoctoral Fellowship. M.T.C. was supported by the Rockefeller University Women and Science Fellowship. R.B.D. is an investigator of the Howard Hughes Medical Institute.

Received: August 2, 2014

Revised: November 26, 2014

Accepted: January 30, 2015

Published: March 12, 2015

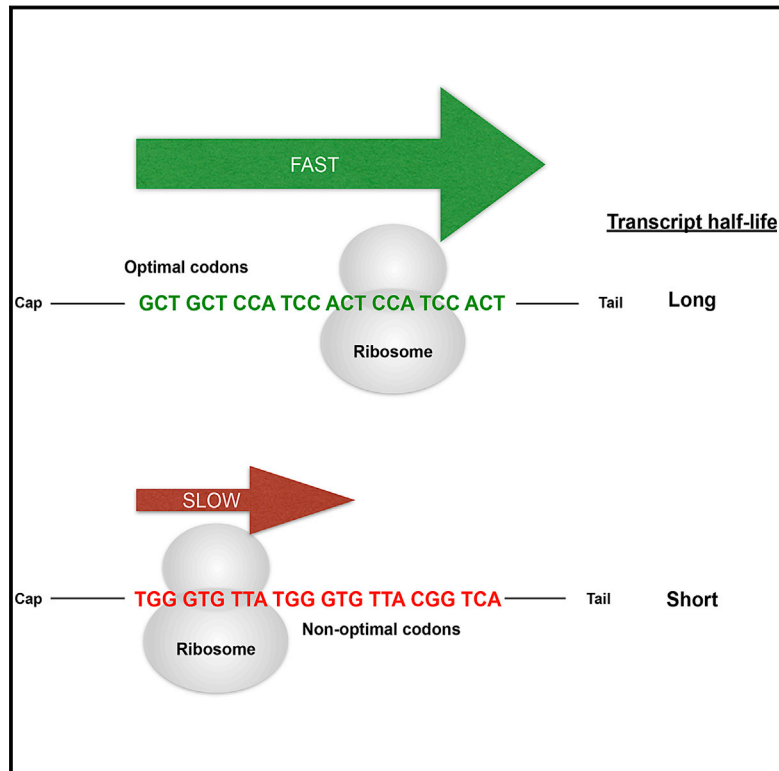
### REFERENCES

- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.
- Castoldi, M., Vujic Spasic, M., Altamura, S., Elmén, J., Lindow, M., Kiss, J., Stolte, J., Sparla, R., D'Alessandro, L.A.L., Klingmüller, U., et al. (2011). The liver-specific microRNA miR-122 controls systemic iron homeostasis in mice. *J. Clin. Invest.* 121, 1386–1396.
- Catanese, M.T., Loureiro, J., Jones, C.T., Dorner, M., von Hahn, T., and Rice, C.M. (2013). Different requirements for scavenger receptor class B type I in hepatitis C virus cell-free versus cell-to-cell transmission. *J. Virol.* 87, 8282–8293.
- Cazalla, D., Yario, T., and Steitz, J.A. (2010). Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* 328, 1563–1566.
- Chang, J., Nicolas, E., Marks, D., Sander, C., Lerro, A., Buendia, M.A., Xu, C., Mason, W.S., Moloshok, T., Bort, R., et al. (2004). miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1. *RNA Biol.* 1, 106–113.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HiTS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486.
- Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261.
- Denzler, R., Agarwal, V., Stefano, J., Bartel, D.P., and Stoffel, M. (2014). Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol. Cell* 54, 766–776.
- Esau, C., Davis, S., Murray, S.F., Yu, X.X., Pandey, S.K., Pear, M., Watts, L., Booten, S.L., Graham, M., McKay, R., et al. (2006). miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.* 3, 87–98.
- Gatfield, D., Le Martelot, G., Vejnar, C.E., Gerlach, D., Schaad, O., Fleury-Olela, F., Ruskeepää, A.-L., Oresic, M., Esau, C.C., Zdobnov, E.M., and Schibler, U. (2009). Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes Dev.* 23, 1313–1326.
- Haecker, I., Gay, L.A.L., Yang, Y., Hu, J., Morse, A.M.A., McIntyre, L.M.L., and Renne, R. (2012). Ago HiTS-CLIP expands understanding of Kaposi's sarcoma-associated herpesvirus miRNA function in primary effusion lymphomas. *PLoS Pathog.* 8, e1002884–e1002884.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C.A., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Henke, J.I., Goergen, D., Zheng, J., Song, Y., Schüttler, C.G., Fehr, C., Jünnemann, C., and Niepmann, M. (2008). microRNA-122 stimulates translation of hepatitis C virus RNA. *EMBO J.* 27, 3300–3310.
- Horwitz, J.A., Dorner, M., Friling, T., Donovan, B.M., Vogt, A., Loureiro, J., Oh, T., Rice, C.M., and Ploss, A. (2013). Expression of heterologous proteins

- flanked by NS3-4A cleavage sites within the hepatitis C virus polyprotein. *Virology* 439, 23–33.
- Hsu, S.-H., Wang, B., Kota, J., Yu, J., Costinean, S., Kutay, H., Yu, L., Bai, S., La Perle, K., Chivukula, R.R., et al. (2012). Essential metabolic, anti-inflammatory, and anti-tumorigenic functions of miR-122 in liver. *J. Clin. Invest.* 122, 2871–2883.
- Janssen, H.L.A., Reesink, H.W., Lawitz, E.J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A.J., Patack, A.K., Chen, A., Zhou, Y., et al. (2013). Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* 368, 1685–1694.
- Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., and Sarnow, P. (2005). Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science* 309, 1577–1581.
- Jopling, C.L., Schütz, S., and Sarnow, P. (2008). Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome. *Cell Host Microbe* 4, 77–85.
- Jung, C.J.C., Iyengar, S., Blahnik, K.R.K., Ajuha, T.P.T., Jiang, J.X.J., Farnham, P.J.P., and Zern, M. (2011). Epigenetic modulation of miR-122 facilitates human embryonic stem cell self-renewal and hepatocellular carcinoma proliferation. *PLoS ONE* 6, e27740–e27740.
- Kandathil, A.J., Graw, F., Quinn, J., Hwang, H.S., Torbenson, M., Perelson, A.S., Ray, S.C., Thomas, D.L., Ribeiro, R.M., and Balagopal, A. (2013). Gastroenterology 145, 1404–1413.e1410.
- Lanford, R.E., Hildebrandt-Eriksen, E.S., Petri, A., Persson, R., Lindow, M., Munk, M.E., Kauppinen, S., and Ørum, H. (2010). Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* 327, 198–201.
- Lee, S., Song, J., Kim, S., Kim, J., Hong, Y., Kim, Y., Kim, D., Baek, D., and Ahn, K. (2013). Selective degradation of host MicroRNAs by an intergenic HCMV noncoding RNA accelerates virus production. *Cell Host Microbe* 13, 678–690.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.
- Li, Y.-P., Gottwein, J.M., Scheel, T.K., Jensen, T.B., and Bukh, J. (2011). MicroRNA-122 antagonism against hepatitis C virus genotypes 1–6 and reduced efficacy by host RNA insertion or mutations in the HCV 5' UTR. *Proc. Natl. Acad. Sci. USA* 108, 4991–4996.
- Li, J., Ghazwani, M., Zhang, Y., Lu, J., Li, J., Fan, J., Gandhi, C.R., and Li, S. (2013a). miR-122 regulates collagen production via targeting hepatic stellate cells and suppressing P4HA1 expression. *J. Hepatol.* 58, 522–528.
- Li, Y., Masaki, T., Yamane, D., McGivern, D.R., and Lemon, S.M. (2013b). Competing and noncompeting activities of miR-122 and the 5' exonuclease Xrn1 in regulation of hepatitis C virus replication. *Proc. Natl. Acad. Sci. USA* 110, 1881–1886.
- Liang, Y., Shilagard, T., Xiao, S.Y., Snyder, N., Lau, D., Cicalese, L., Weiss, H., Vargas, G., and Lemon, S.M. (2009). Visualizing hepatitis C virus infections in human liver by two-photon microscopy. *Gastroenterology* 137, 1448–1458.
- Licatalosi, D.D., Yano, M., Fak, J.J., Mele, A., Grabinski, S.E., Zhang, C., and Darnell, R.B. (2012). Ptpb2 represses adult-specific splicing to regulate the generation of neuronal precursors in the embryonic brain. *Genes Dev.* 26, 1626–1642.
- Machlin, E.S., Sarnow, P., and Sagan, S.M. (2011). Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex. *Proc. Natl. Acad. Sci. USA* 108, 3193–3198.
- Marukian, S., Jones, C.T., Andrus, L., Evans, M.J., Ritola, K.D., Charles, E.D., Rice, C.M., and Dustin, L.B. (2008). Cell culture-produced hepatitis C virus does not infect peripheral blood mononuclear cells. *Hepatology* 48, 1843–1850.
- Mas, V.R., Maluf, D.G., Archer, K.J., Yanek, K., Kong, X., Kulik, L., Freise, C.E., Olthoff, K.M., Ghobrial, R.M., McIver, P., and Fisher, R. (2009). Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. *Mol. Med.* 15, 85–94.
- Moore, M.J., Zhang, C., Gantman, E.C., Mele, A., Darnell, J.C., and Darnell, R.B. (2014). Mapping Argonaute and conventional RNA-binding protein interactions with RNA at single-nucleotide resolution using HITS-CLIP and CIMS analysis. *Nat. Protoc.* 9, 263–293.
- Mukherji, S., Ebert, M.S., Zheng, G.X.Y., Tsang, J.S., Sharp, P.A., and van Oudenaarden, A. (2011). MicroRNAs can generate thresholds in target gene expression. *Nat. Genet.* 43, 854–859.
- Nasheri, N., Singaravelu, R., Goodmurphy, M., Lyn, R.K., and Pezacki, J.P. (2011). Competing roles of microRNA-122 recognition elements in hepatitis C virus RNA. *Virology* 410, 336–344.
- Pal, S., Shuhart, M.C., Thomassen, L., Emerson, S.S., Su, T., Feuerborn, N., Kae, J., and Gretch, D.R. (2006). Intrahepatic hepatitis C virus replication correlates with chronic hepatitis C disease severity in vivo. *J. Virol.* 80, 2280–2290.
- Pang, P.S., Pham, E.A., Elazar, M., Patel, S.G., Eckart, M.R., and Glenn, J.S. (2012). Structural map of a microRNA-122: hepatitis C virus complex. *J. Virol.* 86, 1250–1254.
- Peng, X., Li, Y., Walters, K.-A., Rosenzweig, E.R., Lederer, S.L., Aicher, L.D., Proll, S., and Katze, M.G. (2009). Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers. *BMC Genomics* 10, 373.
- Riley, K.J., Rabinowitz, G.S., Yario, T.A., Luna, J.M., Darnell, R.B., and Steitz, J.A. (2012). EBV and human microRNAs co-target oncogenic and apoptotic viral and human genes during latency. *EMBO J.* 31, 2207–2221.
- Saeed, A., Baloch, K., Brown, R.J.P., Wallis, R., Chen, L., Dexter, L., McClure, C.P., Shakesheff, K., and Thomson, B.J. (2013). Mannan binding lectin-associated serine protease 1 is induced by hepatitis C virus infection and activates human hepatic stellate cells. *Clin. Exp. Immunol.* 174, 265–273.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* 146, 353–358.
- Sedano, C.D., and Sarnow, P. (2014). Hepatitis C virus subverts liver-specific miR-122 to protect the viral genome from exoribonuclease Xrn2. *Cell Host Microbe* 16, 257–264.
- Sheahan, T., Imanaka, N., Marukian, S., Dorner, M., Liu, P., Ploss, A., and Rice, C.M. (2014). Interferon lambda alleles predict innate antiviral immune responses and hepatitis C virus permissiveness. *Cell Host Microbe* 15, 190–202.
- Shimakami, T., Yamane, D., Jangra, R.K., Kempf, B.J., Spaniel, C., Barton, D.J., and Lemon, S.M. (2012). Stabilization of hepatitis C virus RNA by an Ago2-miR-122 complex. *Proc. Natl. Acad. Sci. USA* 109, 941–946.
- Singaravelu, R., Russell, R.S., Tyrrell, D.L., and Pezacki, J.P. (2014). Hepatitis C virus and microRNAs: miRed in a host of possibilities. *Curr Opin Virol* 7, 1–10.
- Skalsky, R.L.R., Corcoran, D.L.D., Gottwein, E., Frank, C.L.C., Kang, D., Hafner, M., Nusbaum, J.D.J., Feederle, R., Delecluse, H.-J.H., Luftig, M.A.M., et al. (2012). The viral and cellular microRNA targetome in lymphoblastoid cell lines. *PLoS Pathog.* 8, e1002484–e1002484.
- Tsai, W.-C., Hsu, S.-D., Hsu, C.-S., Lai, T.-C., Chen, S.-J., Shen, R., Huang, Y., Chen, H.-C., Lee, C.-H., Tsai, T.-F., et al. (2012). MicroRNA-122 plays a critical role in liver homeostasis and hepatocarcinogenesis. *J. Clin. Invest.* 122, 2884–2897.
- Weyn-Vanhenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152.
- Yamane, D., McGivern, D.R., Masaki, T., and Lemon, S.M. (2013). Liver injury and disease pathogenesis in chronic hepatitis C. *Curr. Top. Microbiol. Immunol.* 369, 263–288.
- Zhang, C., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* 29, 607–614.
- Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 40, e54–e54.

# Codon Optimality Is a Major Determinant of mRNA Stability

## Graphical Abstract



## Authors

Vladimir Presnyak, Najwa Alhusaini, ..., Brenton R. Graveley, Jeff Collier

## Correspondence

jmc71@case.edu

## In Brief

Codon usage impacts gene expression both at the level of translation and mRNA decay, with the balance between optimal and non-optimal codons helping to fine-tune levels of mRNAs and, ultimately, proteins.

## Highlights

- Codon identity correlates with yeast mRNA half-lives transcriptome wide
- Converting non-optimal codons to optimal codons increases mRNA stability
- Codon optimality impacts translational elongation rate
- Proteins with related function are coordinated at the level of optimal codon content

## Accession Numbers

GSE57385





# Codon Optimality Is a Major Determinant of mRNA Stability

Vladimir Presnyak,<sup>1,5</sup> Najwa Alhusaini,<sup>1,5</sup> Ying-Hsin Chen,<sup>1,5</sup> Sophie Martin,<sup>1</sup> Nathan Morris,<sup>2</sup> Nicholas Kline,<sup>1</sup> Sara Olson,<sup>4</sup> David Weinberg,<sup>3</sup> Kristian E. Baker,<sup>1</sup> Brenton R. Graveley,<sup>4</sup> and Jeff Collier<sup>1,\*</sup>

<sup>1</sup>Center for RNA Molecular Biology, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>2</sup>Statistical Science Core in the Center for Clinical Investigation, Case Western Reserve University, Cleveland, OH 44106, USA

<sup>3</sup>Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, 94158, USA

<sup>4</sup>Department of Genetics and Developmental Biology, Institute for Systems Genomics, University of Connecticut Health Center, Farmington, CT 06030, USA

<sup>5</sup>Co-first author

\*Correspondence: [jmc71@case.edu](mailto:jmc71@case.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.029>

## SUMMARY

mRNA degradation represents a critical regulated step in gene expression. Although the major pathways in turnover have been identified, accounting for disparate half-lives has been elusive. We show that codon optimality is one feature that contributes greatly to mRNA stability. Genome-wide RNA decay analysis revealed that stable mRNAs are enriched in codons designated optimal, whereas unstable mRNAs contain predominately non-optimal codons. Substitution of optimal codons with synonymous, non-optimal codons results in dramatic mRNA destabilization, whereas the converse substitution significantly increases stability. Further, we demonstrate that codon optimality impacts ribosome translocation, connecting the processes of translation elongation and decay through codon optimality. Finally, we show that optimal codon content accounts for the similar stabilities observed in mRNAs encoding proteins with coordinated physiological function. This work demonstrates that codon optimization exists as a mechanism to finely tune levels of mRNAs and, ultimately, proteins.

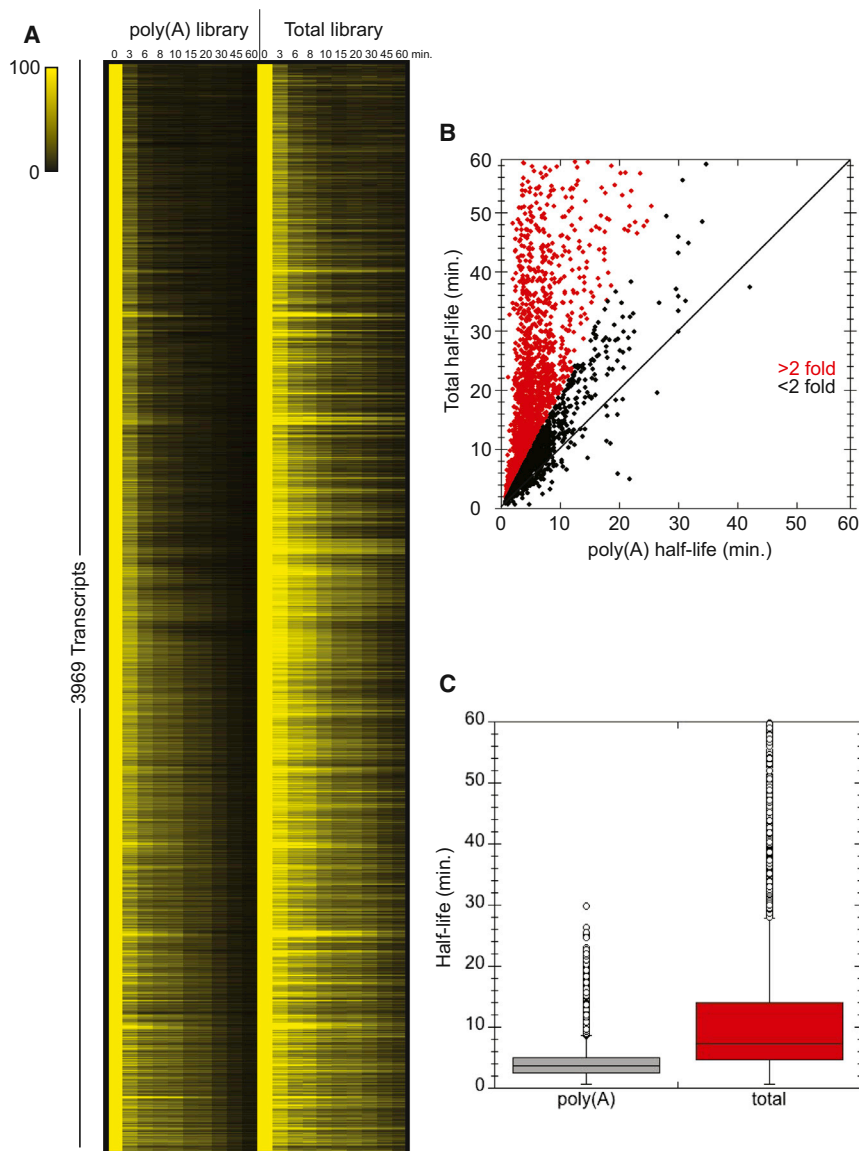
## INTRODUCTION

Messenger RNA (mRNA) degradation plays a critical role in regulating transcript levels in the cell and is a major control point for modulating gene expression. Degradation of most mRNAs in *Saccharomyces cerevisiae* is initiated by removal of the 3' poly(A) tail (deadenylation), followed by cleavage of the 5' 7<sup>m</sup>GpppN cap (decapping) and exonucleolytic degradation of the mRNA body in a 5'-3' direction (Collier and Parker, 2004; Ghosh and Jacobson, 2010). Despite being targeted by a common decay pathway, turnover rates for individual yeast mRNAs differ dramatically, with half-lives ranging from <1 min to 60 min or greater (Collier and Parker, 2004). RNA features that influence transcript stability have long been sought, and some sequence

and/or structural elements located within 5' and 3' UTRs have been implicated in contributing to the decay of a subset of mRNAs (Lee and Lykke-Andersen, 2013; Muhirad and Parker, 1992; Geisberg et al., 2014). However, these features regulate mRNA stability predominantly in a transcript-specific manner through binding of regulatory factors and cannot account for the wide variation in half-lives observed across the entire transcriptome (Geisberg et al., 2014). Therefore, it seems likely that additional and more general features that act to modulate transcript stability could exist within mRNAs.

We have previously shown that inclusion of a cluster of rare arginine codons within the open reading frame (ORF) of a reporter mRNA dramatically enhanced its turnover (Hu et al., 2009; Sweet et al., 2012). The mRNA destabilization caused by rare codons was manifest in the enhancement of both deadenylation and decapping of the transcript. This effect was not dependent on RNA surveillance pathways such as No-Go, Nonsense-Mediated, or Non-Stop Decay (Shoemaker and Green, 2012). The link between rare codons and enhanced mRNA turnover rates of reporter mRNAs is consistent with earlier observations for several endogenous transcripts in yeast (Caponigro et al., 1993; Hoekema et al., 1987).

The rare codons used in our previous studies belong to a general class of codons defined as non-optimal (Pechmann and Frydman, 2013; dos Reis et al., 2004). Conceptually, codon optimality is a scale that reflects the balance between the supply of charged tRNA molecules in the cytoplasmic pool and the demand of tRNA usage by translating ribosomes, representing a measure of translation efficiency. Critically, optimal codons are postulated to be decoded faster and more accurately by the ribosome than non-optimal codons (Akashi, 1994; Drummond and Wilke, 2008), which are hypothesized to slow translation elongation (Novoa and Ribas de Pouplana, 2012; Tuller et al., 2010). Therefore, codon optimality is hypothesized to play an important role in modulation of translation elongation rates and the kinetics of protein synthesis (Krisko et al., 2014; Novoa and Ribas de Pouplana, 2012; Pechmann and Frydman, 2013; dos Reis et al., 2004). In this work, we present four lines of evidence in support of the finding that codon optimality has a broad and powerful influence on mRNA stability in yeast cells. First, global analysis of RNA decay rates reveals that mRNA half-life



**Figure 1. Half-Lives Calculated from Poly(A)<sup>+</sup> versus Total mRNA Differ Significantly**

RNA-seq was performed on poly(A)<sup>+</sup> and total RNA libraries prepared from *rpb1-1* transcriptional shut-off experiments across a 60 min time course. (A) All mRNAs with reliable half-lives in both libraries are plotted visually. Color intensity represents normalized mRNA remaining (time 0 is set to 100% for each mRNA).

(B) Half-life of each mRNA plotted as calculated from total mRNA sequencing against the poly(A) sequencing. Data points with a >2-fold difference are highlighted in red.

(C) Overview of the distribution of half-lives for both libraries.

See also Table S1.

observed similarity in mRNA decay rates for these gene families. Taken together, our data suggest that there is evolutionary pressure on protein-coding regions to coordinate gene expression at the level of protein synthesis and mRNA decay.

## RESULTS

Measuring global mRNA decay rates using methods that either enrich for poly(A)<sup>+</sup> RNA from total RNA samples and/or synthesize complementary DNA (cDNA) using oligonucleotides annealed to the poly(A) tail may fail to capture important information for several reasons. Although it is firmly established that deadenylation is the rate-limiting step in mRNA turnover, we and others have observed that specific mRNAs persist in cells as “stable” deadenylated species (Hu et al., 2009; Muhrad et al., 1995). For such transcripts, decapping and subsequent decay are delayed, and decapping be-

comes the rate-defining step for mRNA degradation. Moreover, some mRNAs may contain structures that impede poly(A) tail function (Geisberg et al., 2014). Lastly, because the process of deadenylation converts an mRNA species from one that can be efficiently captured by oligo dT to one that cannot, the overall level of information gained may vary with the level of poly(A) enrichment achieved in the protocol used. With this in mind, we sought to determine how prevalent these phenomena are on a transcriptome-wide level. For this purpose, we performed a time course after inactivation of RNA polymerase II (Nonet et al., 1987). At each time point, libraries were prepared from either oligo dT-selected mRNAs or rRNA-depleted whole-cell RNA and subjected to Illumina sequencing (see Experimental Procedures). This approach allowed us to compare poly(A) half-lives (oligo dT) with total mRNA decay rates (rRNA depleted; Figure 1A). Remarkably, the vast majority (92%) of transcripts for

correlates with optimal codon content. Many stable mRNAs demonstrate a strong preference toward the inclusion of optimal codons within their coding regions, whereas many unstable mRNAs harbor non-optimal codons. Second, we demonstrate that substitution of optimal codons with synonymous, non-optimal codons results in a dramatic destabilization of the mRNA and that the converse replacement leads to a significant increase in mRNA stability. Third, we experimentally demonstrate an impact of codon optimality on ribosome translocation, indicating that the effect on mRNA decay occurs through modulation of mRNA translation elongation. These findings indicate that transcript-specific translation elongation rate, as dictated by codon usage, is an important determinant of mRNA stability. Fourth, we observe tightly coordinated optimal codon content in genes encoding proteins with common physiological function. We hypothesize that this finding explains the previously

which we could confidently calculate half-lives (3969) had longer half-lives when the rRNA depleted libraries were analyzed relative to the half-lives determined from poly(A)-selected libraries (Figures 1B and 1C). It is important to note that not all of these transcripts exist as deadenylated RNAs because mRNAs with short poly(A) tails will not bind oligo dT. These data indicate that mRNA half-lives determined by oligo dT selection give highly skewed values. For example, the *ADH1* mRNA has a calculated half-life of 4.2 min when determined from poly(A)-selected RNA and a 31.7 min half-life when determined from rRNA-depleted RNA (see Table S1 for complete list).

With this data in hand, we attempted to identify sequence motifs that might dictate stability or instability, without success. Following up on previous observations that inclusion of ten consecutive rare codons in the ORF of an otherwise stable mRNA caused a dramatic decrease in stability (Hu et al., 2009; Sweet et al., 2012), we inspected our transcriptome-wide mRNA half-life data to determine whether codon content within ORFs could affect mRNA stability. To do so, we determined whether mRNAs enriched in any individual codon demonstrated greater or lesser stability. We defined mRNAs as stable if they have a half-life greater than 2-fold longer than the average (~20 min) and unstable if they have a half-life less than half of the average (~5 min). For each codon, we calculated a correlation between the frequency of occurrence of that codon in mRNAs and the stabilities of the mRNAs. Occurrences of a codon were compared to the half-life for each mRNA, and a Pearson correlation calculation was used to generate an R value (graphically represented for sample codons in Figure S1E). We refer to this metric as the codon occurrence to mRNA stability correlation coefficient (CSC). The CSC values for all codons were then compared to each other (Figure 2A). Strikingly, it was observed that some codons preferentially occurred in stable mRNAs, whereas others occurred preferentially in unstable mRNAs (overall p value =  $1.496 \times 10^{-14}$ , permutation p value <  $10^{-4}$ ). For example, the GCT alanine codon was highly enriched in stable transcripts as defined by our RNA-seq analysis, whereas its synonymous codons, GCG and GCA, were preferentially present in unstable transcripts (Figure 2A). Approximately one-third of all codon triplets were over-represented in stable mRNAs, whereas the remaining two-thirds appeared to predominate in unstable mRNAs. As a consequence of the large data set and significance of the observed correlation, these data strongly suggest that codon usage influences mRNA degradation rates.

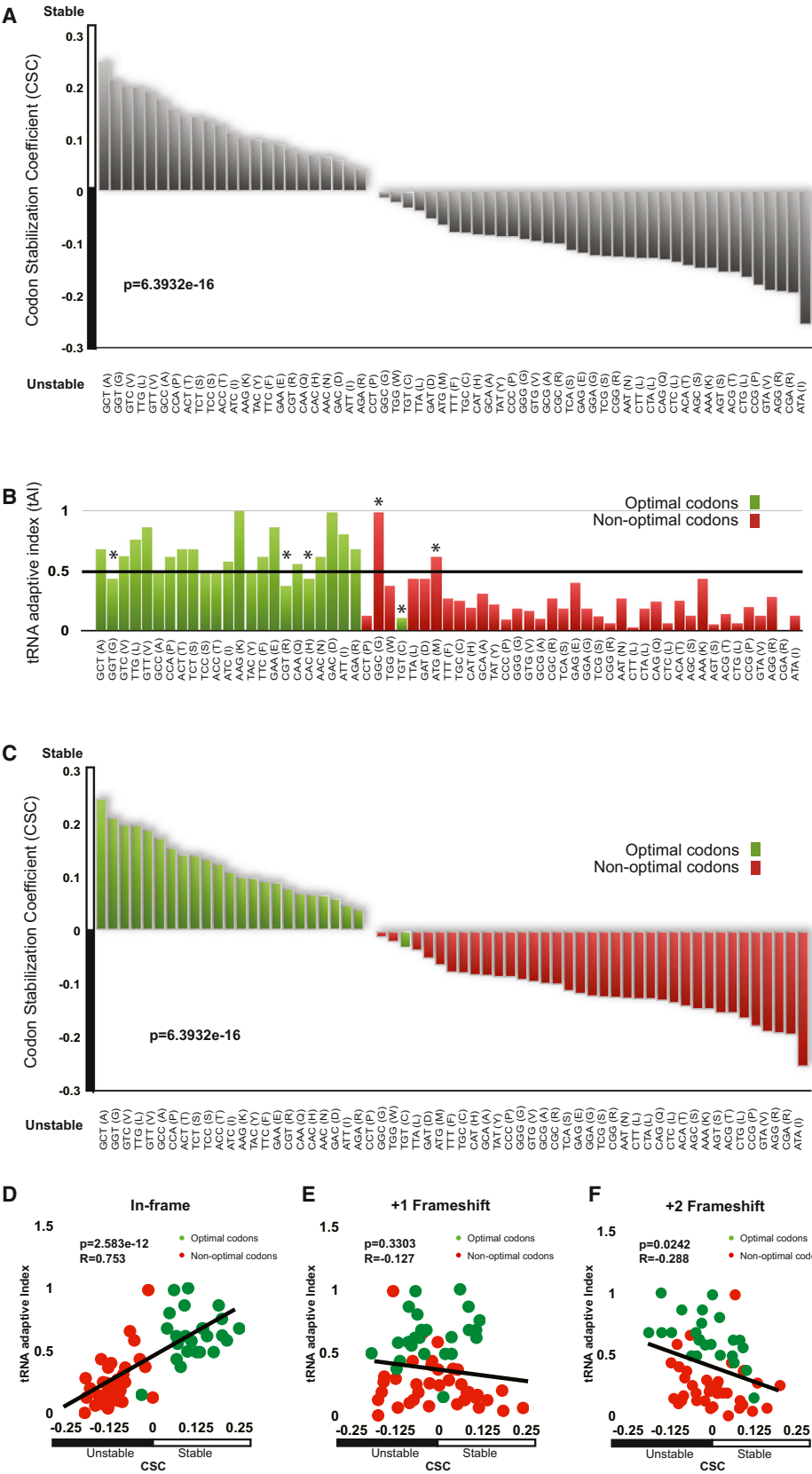
For decades, a large body of literature has hypothesized that some codons may be translated more efficiently than others. dos Reis et al., 2004 laid out a measure of how efficiently a codon would be translated and termed it the tRNA Adaptive Index (tAI). This metric is meant to reflect the efficiency of tRNA usage by the ribosome. The term codon optimality has been introduced in an attempt to define the differential recognition of codons by the translational apparatus (Zhou et al., 2009; Pechmann and Frydman, 2013). Frydman and colleagues generally defined any codon with a tAI above 0.47 as optimal and any codon with a tAI below 0.47 as non-optimal (Pechmann and Frydman, 2013; Figure 2B). Their final designation of codons also takes into account the over- and under-representation of certain codons in

the genome, known as codon bias (Figure 2B, marked with an asterisk [\*]; Zhou et al., 2009; Pechmann and Frydman, 2013). As such, codon optimality is somewhat reflected in genomic codon usage (Figure S1A); however, commonly occurring codons can be optimal or non-optimal, whereas uncommon codons can also be optimal or non-optimal (Figure S1B). Strikingly, codons associated with stable or unstable mRNAs nearly perfectly mirrored their assignment as optimal or non-optimal, respectively (Figure 2C). Direct comparison between our CSC metric and tAI revealed very good overall agreement between these values (Figure 2D;  $R = 0.753$ , p value =  $2.583 \times 10^{-12}$ , permutation p-value <  $10^{-4}$ ). Importantly, the relationship between optimal codon content and mRNA half-life is independent of the method used to determine half-life. We repeated our analysis of codon usage versus mRNA half-life using mRNA decay rates obtained by Miller et al. (2011). In contrast to our own, these data were obtained with a steady-state approach calculation using metabolic labeling that minimally perturbs the cell and is completely distinct from our method (Miller et al., 2011). Both data sets show a similar and striking correlation between optimal codon content and mRNA decay rate (Figures S1C and S1D).

To determine whether the codon optimality correlation was possibly masking other features that might actually be determining mRNA half-life (e.g., sequence content, GC percentage, or secondary structure), we reanalyzed our data after computationally introducing +1 and +2 frameshifts. In the analysis of these frameshifted ORFs, the correlation between codon content and stability completely disappears, thus eliminating other variables as determinative (Figure 2E;  $R = -0.127$ , p value = 0.3303, permutation p value = 0.8847, and Figure 2F;  $R = -0.288$ , p value = 0.0242, permutation p value = 0.0012).

### Stable and Unstable mRNAs Demonstrate Different Optimal Codon Content

As shown above, computational analysis of our global mRNA stability data revealed a relationship between codon occurrence and mRNA half-life. These data indicate that either particular codons alter stability or overall codon content within an mRNA works collectively on stability. To evaluate the relationship between optimal codon content and decay rate on the level of individual transcripts, codon usage was mapped across all individual transcripts (Figure S2). Cluster analysis revealed that different mRNAs are biased toward using different types of codons. The overall result is not surprising, as codon bias has been well studied (Gustafsson et al., 2004); however, the pattern of codon usage demonstrates that certain classes of mRNAs predominately use either optimal or non-optimal codons (Figures 3A and 3B; overrepresented codons in yellow, underrepresented codons in blue) and that this usage correlates with the overall transcript stability (Figure 3C). Closer inspection of several stable mRNAs revealed that these transcripts were not enriched in any particular codon, but an overwhelming proportion (> 80%) of codons fell into the category of optimal (Figure 3D). By contrast, individual unstable mRNAs were found to be enriched (60% or greater) in non-optimal codons (Figure 3E). These analyses demonstrate that, in this set of mRNAs, the stable mRNAs are biased toward harboring predominately optimal codons and the unstable mRNAs are enriched in non-optimal codons,



(legend on next page)



though the specific codon identities vary between individual transcripts.

Extending this analysis to the level of the whole transcriptome, a correlation between optimal codon content and mRNA stability was observed when the proportion of optimal codons within an mRNA was evaluated by percentiles. Specifically, mRNAs with less than 40% optimal codons were typically found to be unstable, with a median half-life of 5.4 min. In contrast, mRNAs with 70% optimal codon content or greater were found to be stable, with a median half-life of 17.8 min (Figure 3F).

### Optimal Codon Content Directly Influences mRNA Decay Rate

To experimentally validate the relationship observed in the computational analysis, we evaluated the effects on stability of altering the percentage of optimal codons within an mRNA. We modified the codon content of the unstable *LSM8* mRNA (half-life = 4.65 min) by making synonymous optimal substitutions in 52 of its 60 non-optimal codons. Similarly, we replaced the majority of optimal codons (108 of 113) within the coding region of the stable *RPS20* mRNA (half-life = 25.3 min) with synonymous, non-optimal codons. This methodology ensured that the polypeptides encoded by these sequences were unchanged from the native form. Moreover, the substitutions were selected to avoid significantly altering the GC content of the coding region or introducing any predicted RNA secondary structure (data not shown). Northern blot analysis of these mRNAs after transcriptional inhibition revealed that alteration of the codons within these two transcripts resulted in dramatic changes in their stability. Specifically, the half-life of *LSM8* mRNA was increased greater than 7-fold as a consequence of the conversion of non-optimal codons into synonymous optimal codons in its ORF (half-life = 18.7 min; Figure 4A). In contrast, substitution of non-optimal for optimal codons within the stable *RPS20* mRNA resulted in a sharp (10-fold) reduction in its stability (half-life = 2.5 min; Figure 4B). These data demonstrate that identity of codons within an mRNA can strongly influence stability and that optimal codon content contributes significantly to determining the rate of mRNA decay in vivo.

To further examine the relationship between optimal codon content and mRNA stability, we generated two synthetic ORFs that encode identical 59 amino acid polypeptides but differ in the optimality at each codon (SYN reporters; Figures S3A–

S3C). We introduced the synthetic ORFs into a reporter bearing the 5' and 3' UTRs of *MFA2*, a well-studied mRNA that is rapidly degraded in the cell (half-life = 3.0 min), a phenomenon shown to be mediated, in part, by elements encoded within its 3' UTR (LaGrande and Parker, 1999; Muhlrud and Parker, 1992). We also introduced the synthetic ORFs into a reporter with the 5' and 3' UTRs of *PGK1*, a well-characterized and stable mRNA (half-life = 25 min; Muhlrud et al., 1995). When stability of the four reporter mRNAs was measured by transcriptional shut-off analysis, the transcripts encoding the optimal SYN ORF were found to be significantly more stable (~4-fold) than their counterparts bearing the non-optimal codons (Figure 4C). Importantly, degradation of both the optimally and non-optimally encoded SYN reporter mRNAs was determined to occur through the deadenylation-dependent decapping pathway used to degrade the majority of endogenous mRNAs in yeast and was not mediated by any of the three pathways known to target aberrant mRNA (Figures S3G and S3H). High-resolution northern analysis of the decay of these mRNAs confirmed that the rates of both deadenylation and decapping, the regulated steps in the normal decay pathway, were affected as a consequence of changes in codon composition within the reporter ORFs (Figures S3D–S3F). These data demonstrate that optimal codon content is a critical determinant of mRNA stability, influencing both the rate of deadenylation and decapping during turnover of the mRNA independently of 5' and 3' UTRs, which can act in parallel to stabilize or destabilize the mRNA.

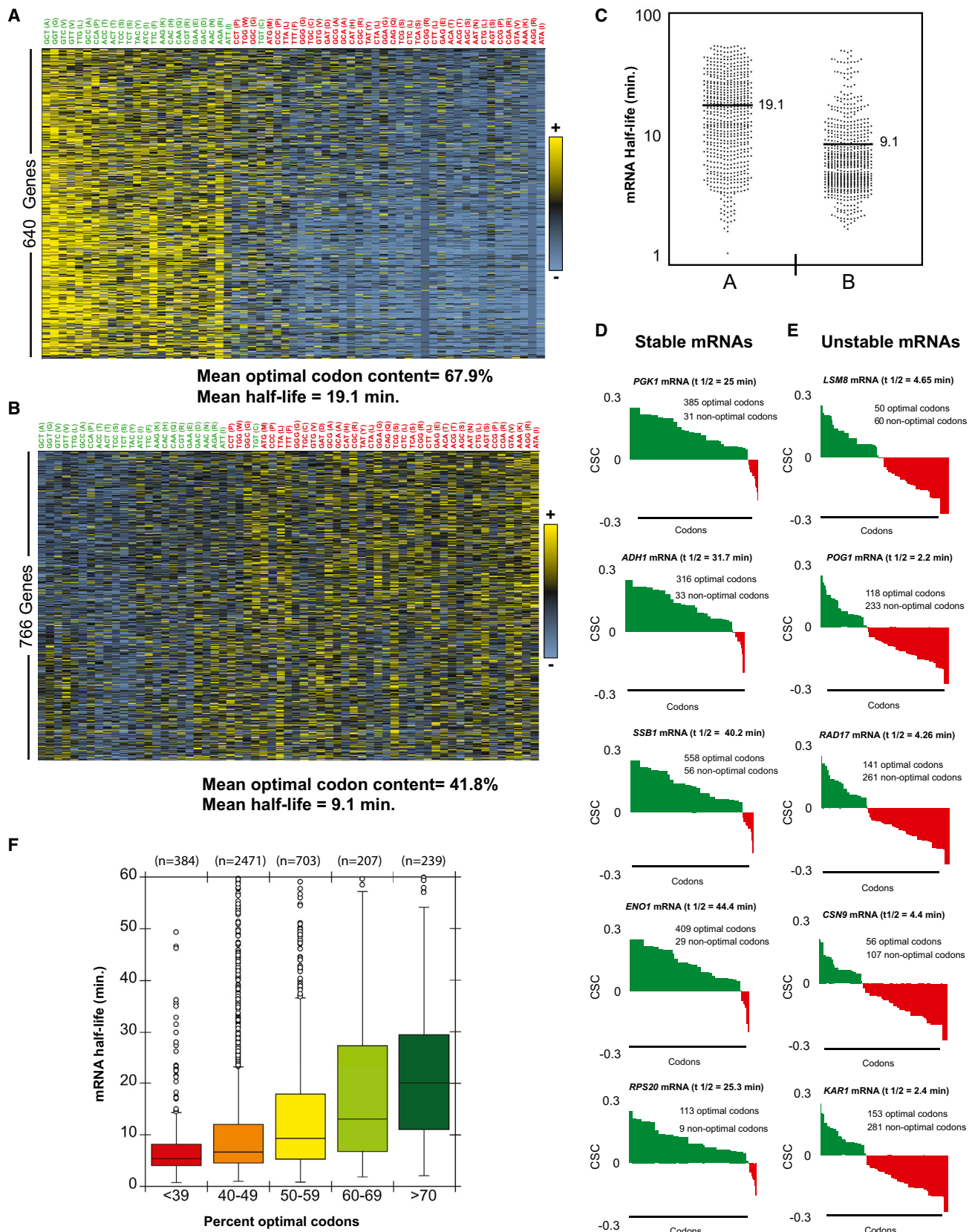
### Optimal Codon Content Influences Translational Efficiency

To evaluate the influence of codon optimality on mRNA translation efficiency in vivo, we generated three new reporters that differ in optimal codon content but do not differ in amino acid sequence. Specifically, we engineered the ORF of the *HIS3* gene to contain either all optimal (*HIS3* opt) or all non-optimal codons (*HIS3* non-opt), with the wild-type *HIS3* gene providing an intermediate point at 43% optimal codons (Figure 5A). The *HIS3* gene was chosen because it has a relatively long ORF (220 amino acids) compared to our other synonymous mutation constructs, allowing us to effectively monitor ribosome association by sucrose density gradients (see below). We then determined the mRNA decay rate of the three *HIS3* constructs by transcriptional shutoff analysis using an *rpb1-1* strain. Consistent with our

### Figure 2. Codon Composition Correlates with Stability

- (A) The CSC plotted for each codon as calculated from the total RNA data set. The CSC is the R value of the correlation between the occurrences of that codon and the half-lives of mRNA. Overall p value is  $6.3932e-16$ , and permutation p value is  $< 10^{-4}$ .
- (B) tRNA adaptive index values for each codon plotted in the same order as (A). Codon optimality as defined in Pechmann and Frydman (2013) is color coded, using green for optimal codons and red for non-optimal codons. Codons designated with an asterisk (\*) were called optimal or non-optimal according to additional criteria discussed therein.
- (C) The CSC plotted for each codon as in (A), but optimality information presented in (B) is added by color-coding. Green color represents optimal codons, and red represents non-optimal codons.
- (D) tRNA adaptive index values plotted versus CSC when ORFs are considered in frame. Green indicates optimal codons, and red indicates non-optimal codons ( $R = 0.7255$ , p value =  $2.075e-09$ , and permutation p value  $< 10^{-4}$ ).
- (E) tRNA adaptive index values plotted versus CSC when ORFs are frameshifted by one nucleotide. Green indicates optimal codons, and red indicates non-optimal codons.
- (F) tRNA adaptive index values plotted versus CSC when ORFs are frameshifted by two nucleotides. Green indicates optimal codons, and red indicates non-optimal codons.

See also Figure S1.



(legend on next page)

previous results, it was observed that changing optimal codon content produced a dramatic effect on mRNA half-life (Figure 5B). Notably, the effect on *HIS3* mRNA decay matched the percent of optimal codons used. The half-life of the optimal construct (half-life > 60 min) was much greater than that of the WT construct (half-life = 9.5 min) whose half-life was markedly greater than the non-optimal construct (half-life = 2.0 min). Thus, we can achieve a full range of mRNA half-lives in yeast without altering protein sequence or flanking sequences by changing optimal codon content.

We hypothesized that codon optimality should influence translation elongation. We tested this hypothesis using two approaches. First, we monitored the protein output from the *HIS3* optimal construct versus the *HIS3* non-optimal construct by western blot and then normalized the protein expression to the mRNA levels, as determined by northern blot. We observed that the non-optimal construct had 4-fold less protein output than the optimal construct (Figure 5C). Second, we evaluated the ribosome density on the *HIS3* mRNA constructs. Ribosome density was monitored using sucrose gradients, followed by fractionation and northern blotting of the isolated fractions (Hu et al., 2009). Critically, it was observed that the ribosome occupancy was nearly identical for all three *HIS3* reporter mRNAs (Figure 5D). Thus, we propose that a 4-fold decrease in protein output, in conjunction with nearly identical localization within a polyribosome, suggests a decrease in ribosome translocation rate on the non-optimal construct as compared to the optimal.

### Optimal Codon Content Impacts Ribosome Translocation

To directly determine whether ribosomes translocate slower on mRNAs containing non-optimal codons versus optimal codons, we monitored ribosomal run-off of these two reporters. To do this, we blocked translational initiation by depriving cells of glucose for 10 min. Glucose deprivation results in rapid inhibition of translational initiation, and thus, bulk polyribosomes are lost by run-off (Coller and Parker, 2005; Figure 6A versus 6C). To monitor ribosomal run-off, we extracted mRNA-ribosome complexes before and after glucose deprivation, separated the material with a sucrose gradient, collected fractions, and monitored the presence of the *HIS3* mRNAs in each fraction by northern analysis. Importantly, under normal conditions, the ribosome occupancy of the *HIS3* opt and non-opt constructs was determined to be similar (Figure 6B); however, upon induction of ribosomal run-off, a large fraction of the optimal construct mRNA relocated

to the top of the gradient in the ribosome-free area, whereas the *HIS3* non-opt mRNA remained largely associated with polyribosomes (Figure 6D). We extended this analysis to two endogenous transcripts that differ dramatically in codon optimality, *LSM8* (45% optimal codons) and *RSP20* (92% optimal codons). Notably, the endogenous *LSM8* mRNA was retained on polyribosomes following inhibition of translational initiation, whereas the *RSP20* mRNA dissociated efficiently. We propose that the difference in retention is due to more efficient ribosome translocation on messages with high optimal codon content. Thus, the retention of the mRNAs bearing predominantly non-optimal codons in polyribosomal fractions indicates that codon optimality can impact the rate of ribosome translocation directly.

### Precision in Gene Expression Is Achieved through Coordination of Optimal Codon Content

A previous analysis of mRNA stability in yeast revealed that the decay rates of some mRNAs encoding proteins that function in the same pathway or are part of the same complex were similar. Turnover of individual mRNAs appears to be based on the physiological function and cellular requirement of the proteins they encode (Wang et al., 2002). We hypothesized that modulation of optimal codon content may provide the mechanism for the cell to coordinate the metabolism of transcripts expressing proteins of common function. We assessed codon usage for genes whose protein products function in common pathways and/or complexes. We observed that mRNAs encoding the enzymes involved in glycolysis ( $n = 10$ ) had a similar and extraordinarily high proportion of optimal codons (mean = 86%; Figure 7A). These transcripts were determined to be stable both previously and in our genome-wide analysis (median half-life = 43.4 min; Wang et al., 2002). In contrast, mRNAs encoding polypeptides involved in pheromone response in yeast cells ( $n = 14$ ) were all unstable (median half-life = 5.6 min; Wang et al., 2002) and harbored an average of only 43% optimal codons (Figure 7A). Our analysis revealed that other groups of transcripts behave similarly. The stable large and small cytosolic ribosomal subunit protein mRNAs ( $n = 70$  and 54, respectively; median half-life = 18.9 min and 20.2 min, respectively) demonstrated an average optimal codon content of 89% and 88%, respectively, but mRNAs that encode ribosomal proteins functioning in the mitochondria are unstable ( $n = 42$ ; median half-life = 4.8 min), which is consistent with the observation that they have 45% optimal codon content (Figures 7A and 7B). Other families of genes that have similar decay rates include those whose protein

### Figure 3. Multiple Codons Are Enriched in Stable and Unstable mRNA Classes

(A) Heat map of a class of relatively stable mRNAs with similar codon usage. Each column represents the usage of a single codon, with each row representing one mRNA. Yellow indicates above average usage of that codon, and blue represents below average usage. See Figure S2 for full heat map.

(B) As in (A), but showing a relatively unstable class of mRNAs.

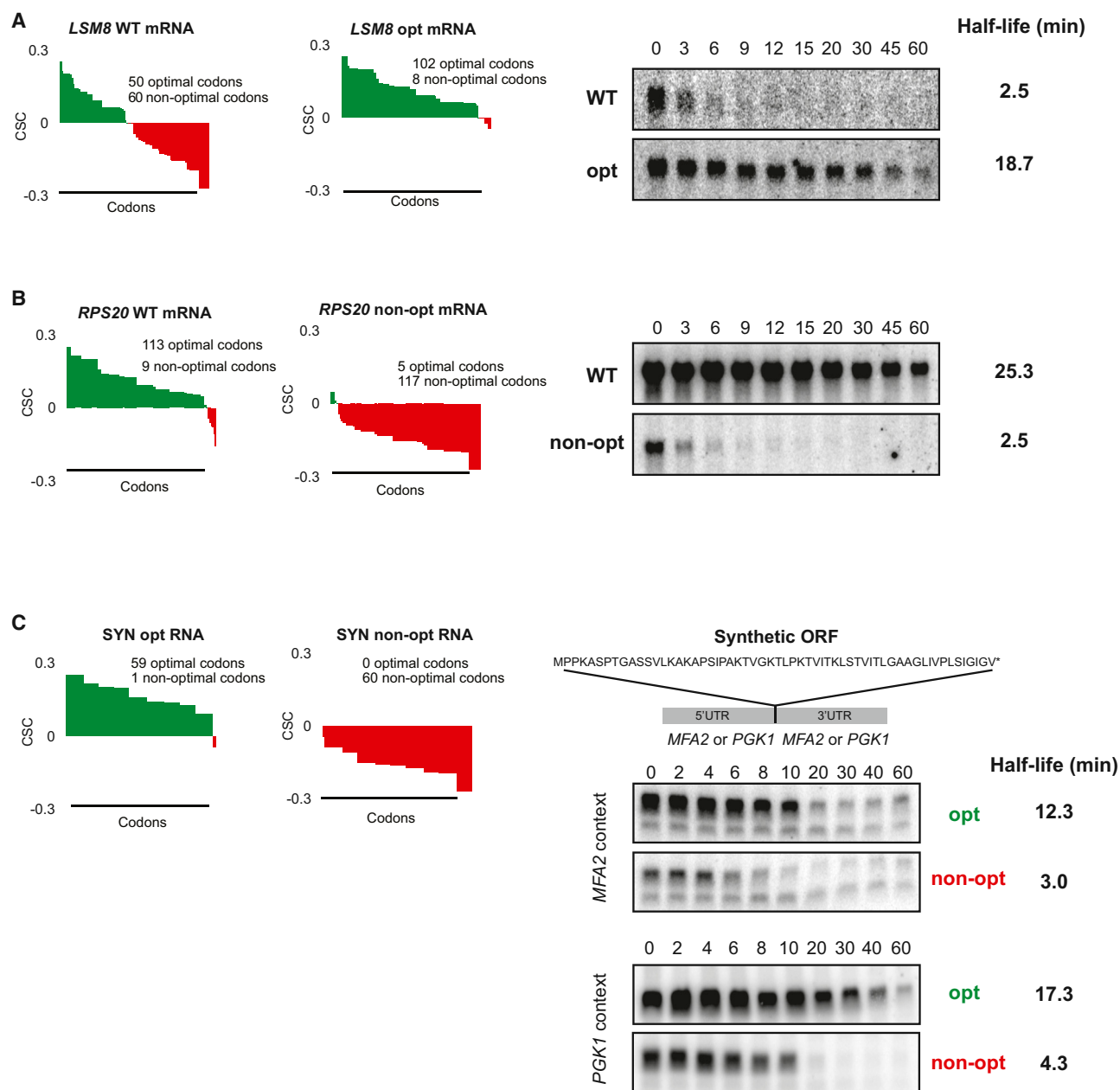
(C) Dot plot showing the distribution of half-lives in the mRNA classes shown in (A) and (B).

(D) Codon optimality diagrams in selected stable mRNAs. Genes are broken down and plotted as individual codons. Codons are presented in order of optimality rather than in their natural order. Higher bars represent more optimal codons (CSC on y axis). Green indicates optimal codons, and red indicates non-optimal codons.

(E) Codon optimality diagrams in selected unstable mRNAs, as in (D).

(F) Box plot of mRNA half-lives separated into optimality groups. Half of the data fall within the boxed section, with the whiskers representing the rest of the data. Data points falling further than 1.5-fold the interquartile distance are considered outliers.

See also Figure S2.



**Figure 4. Stability of mRNAs Can Be Controlled by Altering Codon Optimality**

(A) Codon optimality diagram of *LSM8* (as Figure 3E), a naturally non-optimal mRNA is shown. *LSM8* opt is a synonymously substituted version of *LSM8* engineered for higher optimality. Northern blots of *rpb1-1* shut-off experiments are shown on the right with half-lives of both reporters. Quantitation is normalized to SCR1 loading controls not shown.

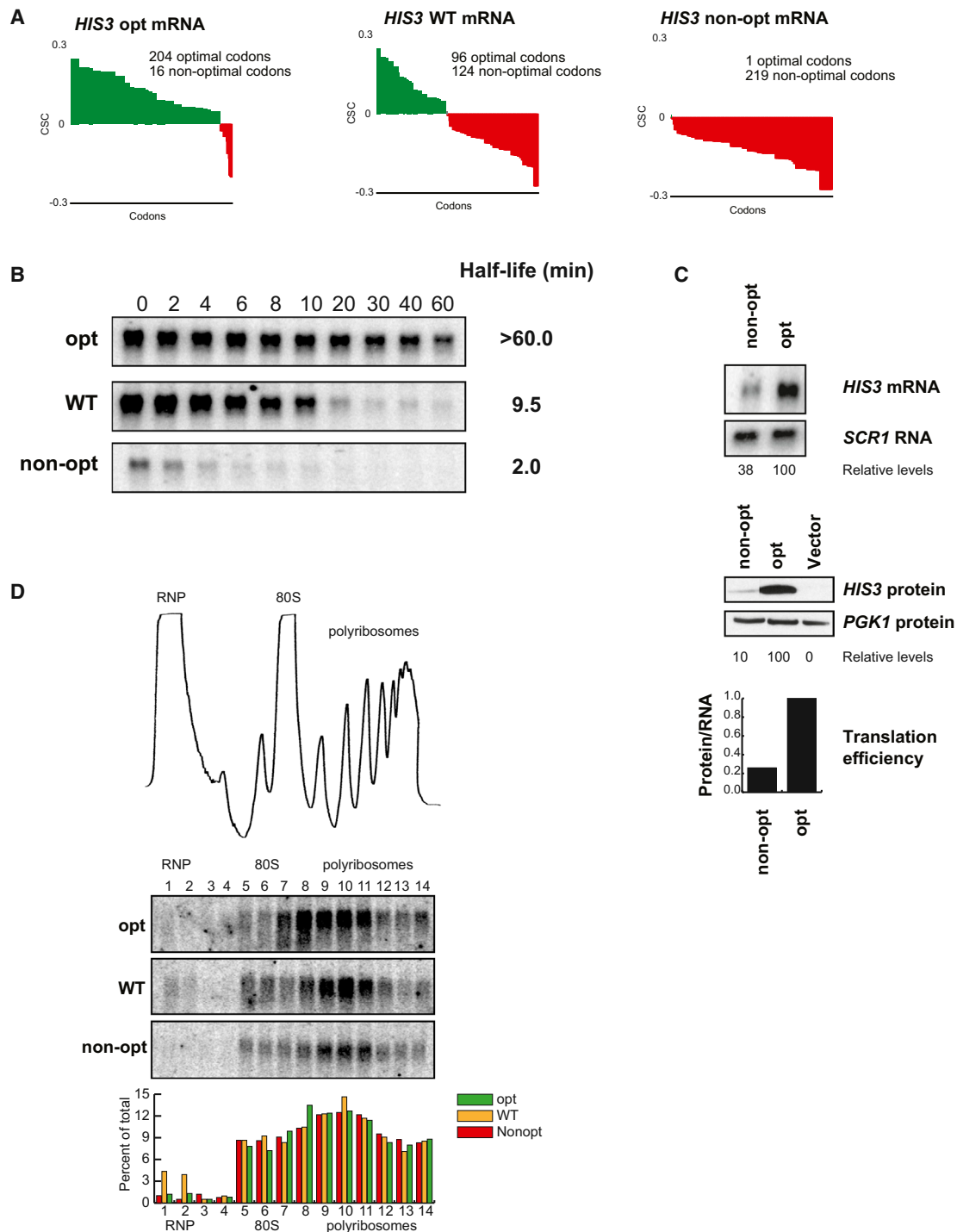
(B) As in (A), except a naturally optimal mRNA, *RPS20* (as in Figure 3D), has been engineered for lower optimality as *RPS20* non-opt. Northern blots of *rpb1-1* shut-off experiments are shown on the right with half-lives of both messages. Quantitation is normalized to SCR1 loading controls not shown.

(C) Codon optimality diagrams showing a synthetic mRNA (*SYN*) encoding the polypeptide shown. Peptide is artificially engineered and has no similarity to any known proteins. *SYN* opt and non-opt were both inserted into flanking regions from a stable transcript (*PGK1*) and unstable transcript (*MFA2*). Northern blots on the right show GAL shut-off experiments demonstrating stability of the *SYN* mRNA in context of the *MFA2* and *PGK1* flanking sequences. Quantitation is normalized to SCR1 loading controls not shown.

See also Figure S3.

products are involved in ribosomal processing, tRNA modification, the TCA cycle, RNA processing, and components of the translational machinery (Figure 7 and data not shown). These data provide evidence that transcripts expressing proteins of related function are coordinated at the level of optimal codon content as well as decay rate, suggesting that these genes





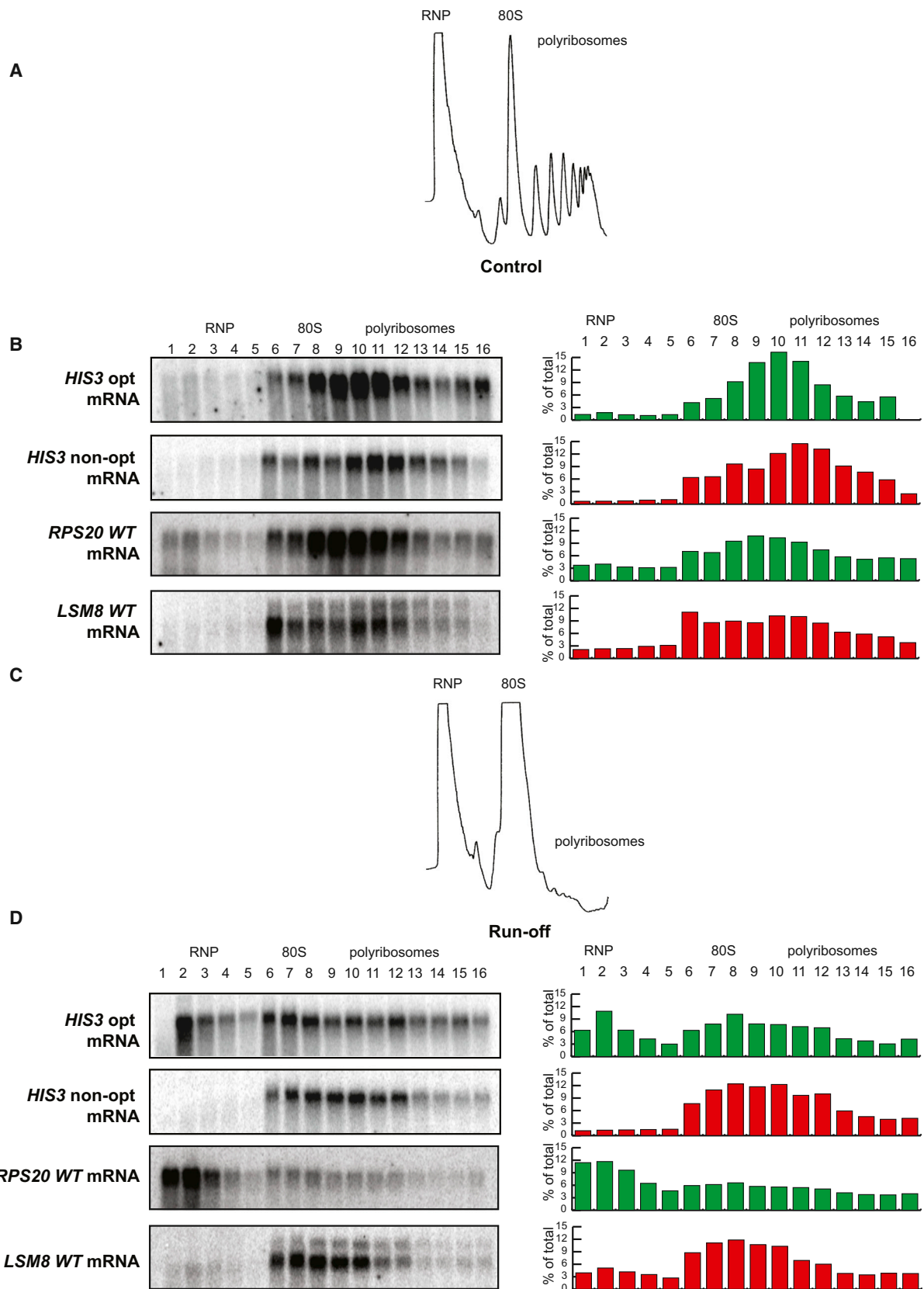
**Figure 5. Optimality Can Affect Translation and Stability of an mRNA without Changes in Ribosome Association**

(A) Codon optimality diagram of *HIS3*, a transcript with an intermediate half-life, as well as versions engineered with synonymous substitutions to contain higher and lower percent optimal codons, *HIS3* opt and *HIS3* non-opt, respectively.

(B) Northern blots of *rpb1-1* shut-off experiments are shown with half-lives of all three messages. Quantitation is normalized to *SCR1*; loading controls not shown.

(C) Northern and Western blots for steady-state concentrations of the optimal and non-optimal versions of *HIS3*. Loading controls and quantitation are shown below. Translational efficiency is calculated as relative protein levels divided by relative mRNA levels and plotted at the bottom. The digital data scans were processed to remove irrelevant lanes from single gel image.

(D) A trace of sucrose density gradient analysis, along with northern blot analysis of the gradient fractions. The blots show location of the three *HIS3* reporters within the gradient. Quantitation for each fraction is shown below.



(legend on next page)

may have evolved specific codon contents as a mechanism to facilitate precise synchronization of expression based on their function in the cell.

## DISCUSSION

We have provided several lines of evidence indicating that codon optimality is a major determinant of mRNA stability in budding yeast. First, bioinformatic analysis demonstrates a strong correlation between the percentage of optimal codons and mRNA half-life. For example, mRNAs with less than 40% optimal codons have a median half-life of 5.3 min, whereas mRNAs with greater than 70% optimal codons have a median half-life of 20.1 min. The conclusions emerging from the bioinformatics were verified experimentally, showing that changing optimal codons to non-optimal destabilized otherwise stable mRNAs while changing non-optimal codons to optimal stabilized otherwise unstable mRNAs. Most importantly, we provide evidence that optimal and non-optimal codons exert their effects by modulating translational elongation rates.

Several ribosomal profiling studies have failed to detect codon-specific differences in the translation of optimal and non-optimal codons (Ingolia et al., 2009; Qian et al., 2012; Charnecki and Hurst, 2013). Nevertheless, we observe striking differences in ribosome clearance when mRNAs encoding the same polypeptide are composed of optimal or non-optimal codons. These differences may reflect the additive effects of many small ribosome hesitations at non-optimal codons. Such hesitations would be imperceptible in ribosomal profiling analyses. Alternatively, the overall codon composition of an ORF could set a uniform translational elongation rate across the ORF. If this were true, no change in rate at individual codons would be detected by ribosome profiling.

It is important to note that, although codon content is clearly a major determinant of mRNA stability, it does not predict half-lives of all mRNAs. For example, mRNAs for several histone components, such as HHF2 and HHT1, contain 85% optimal codons but yet are very unstable with half-lives of 2.4 and 3.5 min, respectively. The half-lives of such mRNAs could be dictated by their ability to initiate translation efficiently (or inefficiently) and/or by elements in 5' or 3' UTRs. Numerous examples of each have been described (Goldstrohm et al., 2007; Olivas and Parker, 2000). It is also possible that features within the ORFs might explain some of the outliers (e.g., distribution of optimal and non-optimal codons).

Because of the effects of optimal codon content on translational elongation rates, it is most likely that some factor(s) monitor these rates while mRNAs are engaged with ribosomes. Indeed, we have previously shown that slowing of ribosomal movement by insertion of rare codons promotes mRNA decay (Hu et al., 2009; Sweet et al., 2012). A prime candidate for a moni-

toring factor is the DEAD-box RNA helicase DHH1, an integral component of the mRNA decay machinery (Presnyak and Collier, 2013) that has been shown to act as an activator of decapping through its role in promoting translational repression (Collier and Parker, 2005; Sweet et al., 2012). Further studies will be needed to determine the mechanism by which translational elongation rate influences mRNA decay.

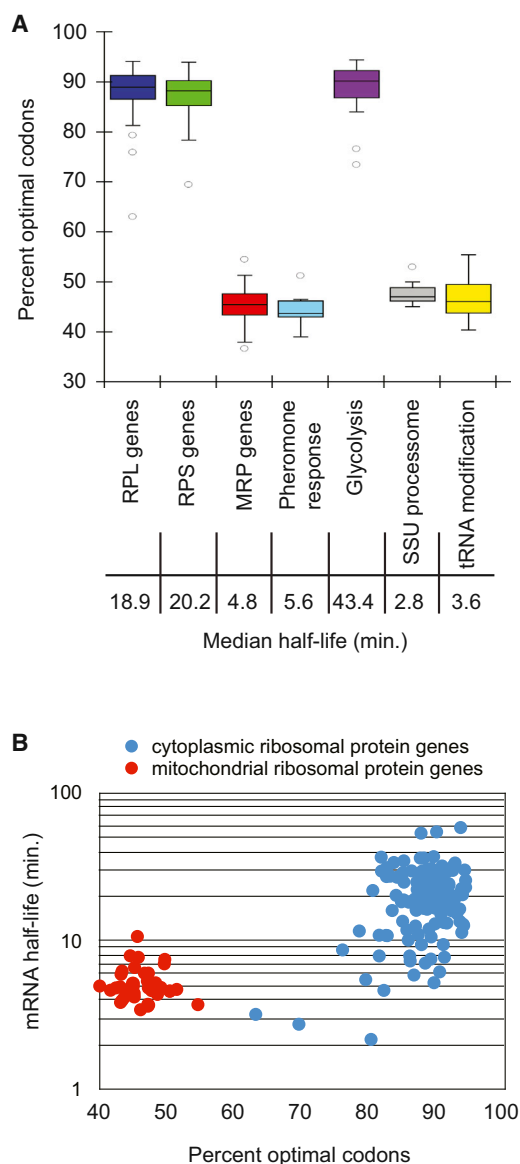
## Precision and Coordination of Gene Expression through Codon Optimality

Both long and short timescales provide important opportunities for the reassignment of codon optimality in the cell. In the short term, changes in cellular growth conditions and nutrient availability could significantly impact individual (or subsets of) charged tRNA levels. As a consequence of this reduction in supply, translational elongation rates of mRNAs enriched in the codons decoded by these tRNAs would be slowed and their levels decreased, due to enhanced turnover. In this way, codon optimality provides the cell not only with a general mechanism to hone mRNA levels but also with a mechanism to sense environmental conditions and rapidly tailor global patterns of gene expression.

Long-term genetic changes can introduce synonymous mutations into protein coding genes that do not alter the amino acid sequence of the encoded polypeptide; however, such changes would impact mRNA and protein expression levels if the mutations significantly altered the proportion of optimal codons within the ORF of the mRNA. Thus, synonymous gene mutation can be envisioned as a method to evolve mRNA stability rates that are advantageous to the cell. We find that mRNAs encoding proteins that act together in similar pathways or are part of the same stoichiometric complexes, and which have been previously observed to decay at similar rates (Wang et al., 2002), encode nearly identical proportions of optimal codons (Figure 7A). We suggest that codon optimality has been finely tuned for these gene sets as an elegant mechanism to ensure coordinated post-transcriptional regulation and parsimonious expression of proteins at the precise levels required by the cell. Interestingly, similar levels of optimal to non-optimal codons could ensure not only similarity of stability and translation rates for related mRNAs but also coordination of response to changes in tRNA levels (e.g., nutrient availability, stress, cell type, etc.). Recent studies reveal that tRNA concentrations within the cell are not static but are constantly undergoing change, sometimes dramatically. For instance, large-scale RNA profiling experiments have demonstrated that tRNA concentrations vary widely between proliferating and differentiating cells (Gingold et al., 2014). Based on our analysis, we would argue that significant alterations in tRNA concentrations could alter the mRNA expression profile within a cell by dynamically changing mRNA stability, even without any changes in transcription.

### Figure 6. Optimal and Non-optimal Transcripts Are Retained Differently on Polysomes

- (A) Representative  $A_{260}$  trace of sucrose density gradient analysis demonstrating normal distribution into RNP, 80S, and polyribosome fractions.  
 (B) Distribution of the optimal and non-optimal *HIS3* reporters and the *RPS20* and *LSM8* mRNAs in the sucrose density gradients under normal conditions showing localization primarily in the polyribosome fractions.  
 (C) Representative  $A_{260}$  trace of sucrose density gradient analysis under run-off conditions, showing collapse of the polyribosome fractions.  
 (D) Distribution of the optimal and non-optimal *HIS3* reporters and the *RPS20* and *LSM8* mRNAs under run-off conditions, demonstrating differential relocation.



**Figure 7. Functionally Related Genes Display Similar Optimality**

(A) Groups of genes whose protein products have related functions are plotted to show their optimality. Half of the data fall within the boxed section, with the whiskers representing the rest of the data. Data points falling further than 1.5-fold the interquartile distance are considered outliers. Represented gene groups are 70 RPL (large ribosomal subunit proteins) genes, 54 RPS (small ribosomal subunit proteins) genes, 42 MRP (mitochondrial ribosomal proteins) genes, 14 pheromone response genes, 10 glycolysis enzymes, 15 SSU (small subunit processome) genes, and 12 tRNA processing genes.

(B) Breakdown of two groups to show relationship between optimal codon content and half-life within the groups. mRNA half-life for each protein in the cytoplasmic ribosome and the mitochondrial ribosome is plotted against the optimal codon content of that mRNA.

### Ribosomes Are the Master Gatekeepers, Determining the Downstream Fate of Both Normal and Aberrant mRNAs

As a final implication, our work suggests that co-translational mRNA surveillance by the ribosome is not only important to

target aberrant mRNAs to rapid decay but also to tune the degradation rates of normal mRNAs. In eukaryotes, aberrations in mRNAs lead to aberrant translation events such as premature termination, lack of translation termination, and ribosome stalling, which result in the accelerated turnover of the mRNA by the Nonsense-Mediated, Non-Stop, and No-Go Decay pathways, respectively (Shoemaker and Green, 2012). We find here that codon usage within normal mRNAs also influences translating ribosomes and can have profound effects on mRNA stability. Thus, the ribosome acts as the master sensor, helping to determine the fate of all mRNAs, both normal and aberrant, through modulation of its elongation and/or termination processes. The use of the ribosome as a sensor is ideal for protein-coding genes, whose primary function in the cell is to be translated. We suggest that a component of mRNA stability is built into all mRNAs as a function of codon composition. The elongation rate of translating ribosomes is communicated to the general decay machinery, which affects the rate of deadenylation and decapping. Individually, the identity of codons within an mRNA would be predicted to have a minute influence on overall ribosomal decoding; however, within the framework of an entire mRNA, we show that codon optimality can have profound effects on translation elongation and mRNA turnover. We therefore conclude that codon identity represents a general property of mRNAs and is a critical determinant of their stability.

### EXPERIMENTAL PROCEDURES

#### Yeast Strains and Growth Conditions

The genotypes of all yeast strains used in this study are listed in Table S2. Unless indicated, all strains are based on BY4741. Cells were grown in standard synthetic medium (pH 6.5) supplemented with appropriate amino acids and sugars. All cells were grown at 24°C and collected at mid-log phase ( $3 \times 10^7$  cells  $\text{ml}^{-1}$ ).

#### Plasmids and Strain Construction

The plasmids and oligonucleotides used in this study are listed in Tables S3 and S4, respectively. Reporter plasmids bearing native genes (*LSM8*, *RPS20*, *HIS3* WT) were constructed by amplifying the native loci, adding restriction sites and several unique sites (to facilitate detection by northern probe) in the 3' UTR by site-directed mutagenesis, and inserting the construct into an expression vector. The reporters with altered optimality (*LSM8* opt, *RPS20* non-opt, *HIS3* opt, and non-opt) were constructed by synthesizing the DNA in multiple pieces, annealing and amplifying them, and then subcloning into an expression vector. These reporter plasmids were transformed into an *rpb1-1* yeast strain.

To construct the plasmids bearing the synthetic reporters, restriction sites were introduced into previously constructed plasmids bearing *MFA2* and *PGK1* under the control of a GAL1 UAS. The SYN ORFs were then synthesized and assembled as described for the altered reporters above. These reporters were transformed into a WT yeast strain.

#### Northern RNA Analysis and Sucrose Density Gradients

Northern RNA analysis of GAL-driven reporters and sucrose density gradients for polyribosome analysis were performed as previously described (Hu et al., 2009). Analysis of reporters in *rpb1-1* was performed similarly to GAL, except cells were grown in media containing glucose and repression was achieved by shifting cells to 37°C. Ribosomal run-off experiments were performed similarly to normal polyribosome analysis, except cells were resuspended in media lacking glucose for 10 min before harvesting (Collier and Parker, 2005).

#### RNA-Seq

*rpb1-1* mutant cells (Nonet et al., 1987) were grown to mid-log phase at 24°C and shifted to a non-permissive temperature of 37°C. Aliquots were collected



over 60 min. RNA was then extracted, external controls were added, and two sets of libraries were prepared from each using the Illumina TruSeq Stranded Total RNA and mRNA library prep kits. The libraries were quantitated using an Agilent Bioanalyzer and sequenced on an Illumina HiSeq2000 using paired-end 100 bp reads with an index read. Sequencing data and the processed data for each gene are available at the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE57385.

### Alignment and Half-Life Calculation

Reads were aligned to the *S. cerevisiae* reference genome using bowtie (Langmead et al., 2009), with the unaligned reads then aligned to the sequences of the controls in the same way. Aligned reads were quantitated using cufflinks (Trapnell et al., 2010). Raw FPKM numbers were normalized to external controls and then fitted to single exponential decay curves to calculate the half-lives using the least absolute deviation method to minimize outlier effects. Data were then filtered to exclude dubious ORFs and transcripts with poor fit to the model. Bootstrapped confidence intervals were generated by using un-normalized residuals from the original data to generate simulated data sets.

### Statistical Methods

The CSC was determined by calculating a Pearson correlation coefficient between the frequency of occurrence of individual codons and the half-lives of the messages containing them. To determine the statistical significance, we categorized the CSC as either positive or negative and used a chi-square test of association. For association between the categories of percent optimal codons and mRNA half-life, an ANOVA f-test with mRNA half-life on the log scale was used.

To mitigate effects of base pair content of the genes, we randomly permuted the sequence and recalculated the test statistic for each of 10,000 permutations. The permutation p value was calculated as the number of permuted data sets with a test of association stronger than the chi-square test in the original data.

Statistical calculations were done using the R environment. Optimality percentages were calculated by generating a list of optimal and non-optimal codons as previously described (Pechmann and Frydman, 2013).

### ACCESSION NUMBERS

Sequencing data and the processed data for each gene are available at the Gene Expression Omnibus under accession number GSE57385.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.029>.

### AUTHOR CONTRIBUTIONS

V.P., K.E.B., B.R.G., and J.C. wrote the manuscript. V.P., Y.-H.C., N.A., S.M., and S.O. performed the experiments. D.W. provided technical expertise, N.K. provided bioinformatic support, and N.M. provided statistical support. All of the authors contributed to discussion, analysis, and the design of the research. All authors commented on the manuscript.

### ACKNOWLEDGMENTS

The authors express their appreciation to Dr. Tim Nilsen for his suggestions and critical evaluation of the manuscript and Dr. Beth Grayhack (University of Rochester) for helpful discussion. We thank the members of the Collier and Baker labs for their input into this work. This work was funded by NIH grant GM095621 to K.E.B. and GM080465 to J.C. K.E.B. and J.C. are inventors on a patent that has been filed related to this work.

Received: August 8, 2014

Revised: November 18, 2014

Accepted: February 11, 2015

Published: March 12, 2015

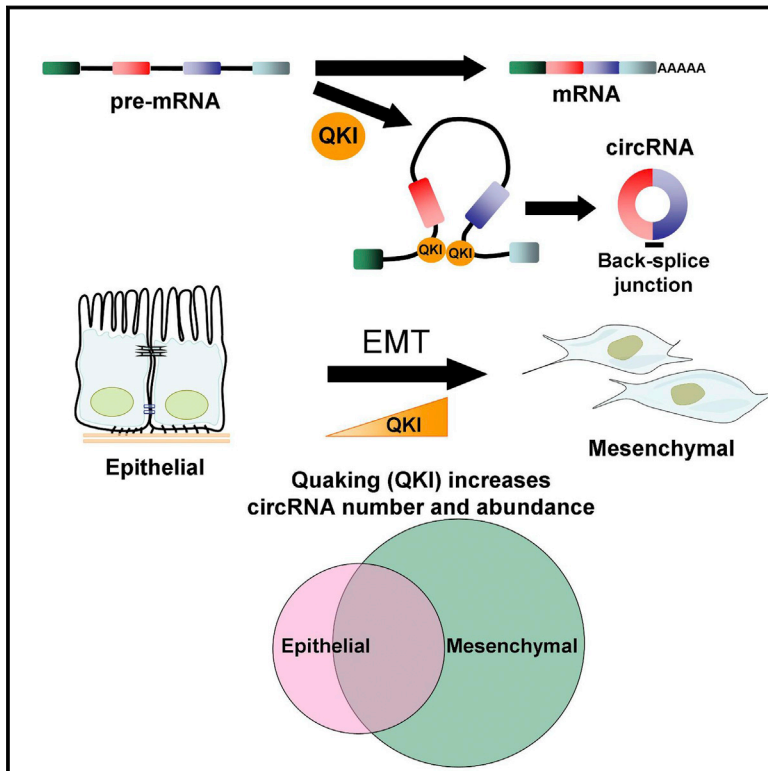
### REFERENCES

- Akashi, H. (1994). Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136, 927–935.
- Caponigro, G., Muhrad, D., and Parker, R. (1993). A small segment of the MAT alpha 1 transcript promotes mRNA decay in *Saccharomyces cerevisiae*: a stimulatory role for rare codons. *Mol. Cell. Biol.* 13, 5141–5148.
- Charneski, C.A., and Hurst, L.D. (2013). Positively charged residues are the major determinants of ribosomal velocity. *PLoS Biol.* 11, e1001508.
- Collier, J., and Parker, R. (2004). Eukaryotic mRNA decapping. *Annu. Rev. Biochem.* 73, 861–890.
- Collier, J., and Parker, R. (2005). General translational repression by activators of mRNA decapping. *Cell* 122, 875–886.
- Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134, 341–352.
- Geisberg, J.V., Moqtaderi, Z., Fan, X., Oszolac, F., and Struhl, K. (2014). Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* 156, 812–824.
- Ghosh, S., and Jacobson, A. (2010). RNA decay modulates gene expression and controls its fidelity. *Wiley Interdiscip. Rev. RNA* 1, 351–361.
- Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158, 1281–1292.
- Goldstrohm, A.C., Seay, D.J., Hook, B.A., and Wickens, M. (2007). PUF protein-mediated deadenylation is catalyzed by Ccr4p. *J. Biol. Chem.* 282, 109–114.
- Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* 22, 346–353.
- Hoekema, A., Kastelein, R.A., Vasser, M., and de Boer, H.A. (1987). Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.* 7, 2914–2924.
- Hu, W., Sweet, T.J., Chamnongpol, S., Baker, K.E., and Collier, J. (2009). Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* 461, 225–229.
- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223.
- Krisko, A., Copic, T., Gabaldón, T., Lehner, B., and Supek, F. (2014). Inferring gene function from evolutionary change in signatures of translation efficiency. *Genome Biol.* 15, R44.
- LaGrandeur, T., and Parker, R. (1999). The cis acting sequences responsible for the differential decay of the unstable MFA2 and stable PGK1 transcripts in yeast include the context of the translational start codon. *RNA* 5, 420–433.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lee, S.R., and Lykke-Andersen, J. (2013). Emerging roles for ribonucleoprotein modification and remodeling in controlling RNA fate. *Trends Cell Biol.* 23, 504–510.
- Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marciniowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol. Syst. Biol.* 7, 458.
- Muhrad, D., and Parker, R. (1992). Mutations affecting stability and deadenylation of the yeast MFA2 transcript. *Genes Dev.* 6, 2100–2111.
- Muhrad, D., Decker, C.J., and Parker, R. (1995). Turnover mechanisms of the stable yeast PGK1 mRNA. *Mol. Cell. Biol.* 15, 2145–2156.

- Nonet, M., Scafe, C., Sexton, J., and Young, R. (1987). Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* 7, 1602–1611.
- Novoa, E.M., and Ribas de Pouplana, L. (2012). Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* 28, 574–581.
- Olivas, W., and Parker, R. (2000). The Puf3 protein is a transcript-specific regulator of mRNA degradation in yeast. *EMBO J.* 19, 6602–6611.
- Pechmann, S., and Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nat. Struct. Mol. Biol.* 20, 237–243.
- Presnyak, V., and Collier, J. (2013). The DHH1/RCKp54 family of helicases: an ancient family of proteins that promote translational silencing. *Biochim. Biophys. Acta* 1829, 817–823.
- Qian, W., Yang, J.R., Pearson, N.M., Maclean, C., and Zhang, J. (2012). Balanced codon usage optimizes eukaryotic translational efficiency. *PLoS Genet.* 8, e1002603.
- dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* 32, 5036–5044.
- Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* 19, 594–601.
- Sweet, T., Kovalak, C., and Collier, J. (2012). The DEAD-box protein Dhh1 promotes decapping by slowing ribosome movement. *PLoS Biol.* 10, e1001342.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., and Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141, 344–354.
- Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA* 99, 5860–5865.
- Zhou, T., Weems, M., and Wilke, C.O. (2009). Translationally optimal codons associate with structurally sensitive sites in proteins. *Mol. Biol. Evol.* 26, 1571–1580.

# The RNA Binding Protein Quaking Regulates Formation of circRNAs

## Graphical Abstract



## Authors

Simon J. Conn, Katherine A. Pillman, ..., Philip A. Gregory, Gregory J. Goodall

## Correspondence

greg.goodall@health.sa.gov.au

## In Brief

The RNA binding protein Quaking (QKI) promotes circRNA biogenesis during epithelial to mesenchymal transition, strongly arguing for their functions.

## Highlights

- Numerous circRNAs are regulated during epithelial to mesenchymal transition
- Quaking (QKI) regulates circRNA formation during EMT
- QKI binds to sites flanking circRNA-forming exons
- Insertion of QKI binding sites into linear RNA can induce exon circularization



# The RNA Binding Protein Quaking Regulates Formation of circRNAs

Simon J. Conn,<sup>1</sup> Katherine A. Pillman,<sup>1,2</sup> John Toubia,<sup>1,2</sup> Vanessa M. Conn,<sup>1</sup> Marika Salamanidis,<sup>1</sup> Caroline A. Phillips,<sup>1,3</sup> Suraya Roslan,<sup>1</sup> Andreas W. Schreiber,<sup>1,2,3</sup> Philip A. Gregory,<sup>1,4</sup> and Gregory J. Goodall<sup>1,3,4,\*</sup>

<sup>1</sup>Centre for Cancer Biology, SA Pathology and University of South Australia, Adelaide, SA 5000, Australia

<sup>2</sup>ACRF Cancer Genomics Facility, SA Pathology, Adelaide, SA 5000, Australia

<sup>3</sup>School of Molecular and Biomedical Science, University of Adelaide, Adelaide, SA 5005, Australia

<sup>4</sup>Department of Medicine, University of Adelaide, Adelaide, SA 5005, Australia

\*Correspondence: [greg.goodall@health.sa.gov.au](mailto:greg.goodall@health.sa.gov.au)

<http://dx.doi.org/10.1016/j.cell.2015.02.014>

## SUMMARY

Circular RNAs (circRNAs), formed by non-sequential back-splicing of pre-mRNA transcripts, are a widespread form of non-coding RNA in animal cells. However, it is unclear whether the majority of circRNAs represent splicing by-products without function or are produced in a regulated manner to carry out specific cellular functions. We show that hundreds of circRNAs are regulated during human epithelial-mesenchymal transition (EMT) and find that the production of over one-third of abundant circRNAs is dynamically regulated by the alternative splicing factor, Quaking (QKI), which itself is regulated during EMT. Furthermore, by modulating QKI levels, we show the effect on circRNA abundance is dependent on intronic QKI binding motifs. Critically, the addition of QKI motifs is sufficient to induce *de novo* circRNA formation from transcripts that are normally linearly spliced. These findings demonstrate circRNAs are both purposefully synthesized and regulated by cell-type specific mechanisms, suggesting they play specific biological roles in EMT.

## INTRODUCTION

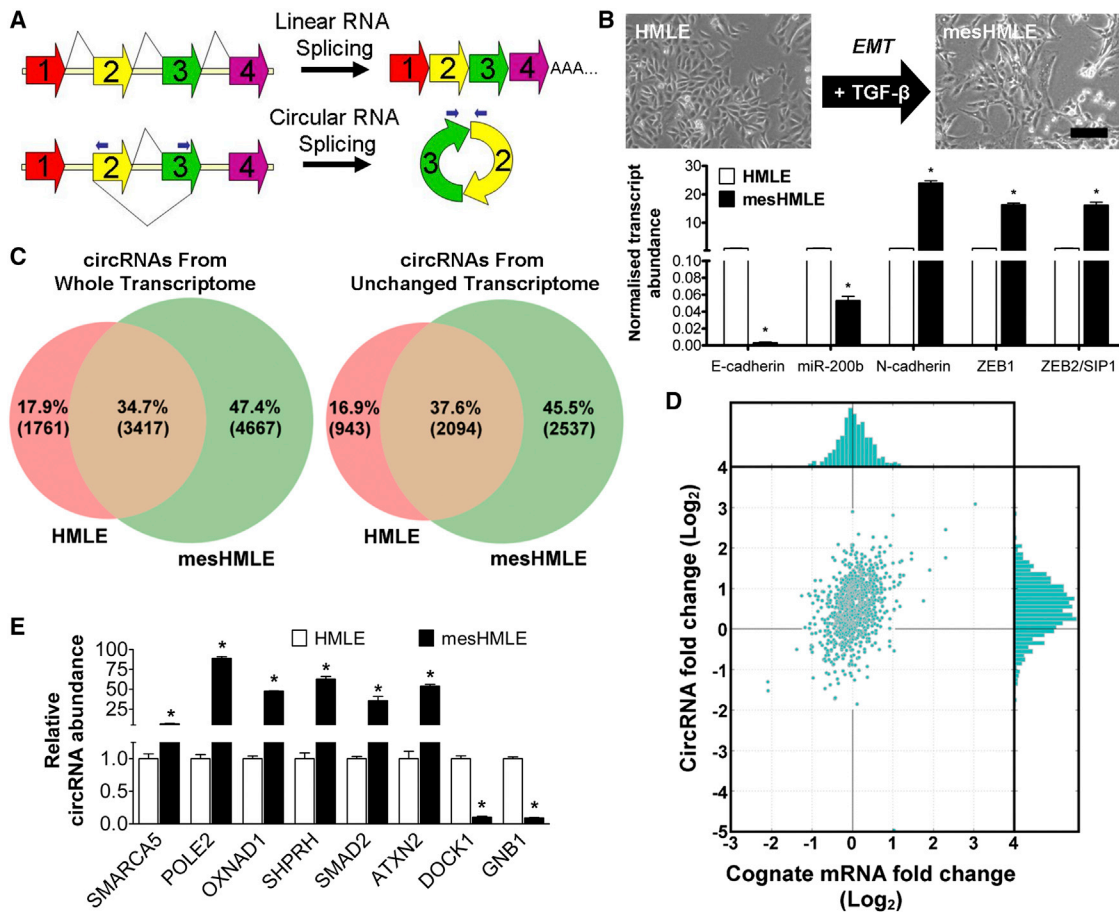
Circular RNAs (circRNAs) were identified in the early 1990s as transcripts with scrambled exon order (Nigro et al., 1991) and continued to be reported for a number of transcripts over the following two decades (Capel et al., 1993; Cocquerelle et al., 1993; Gualandi et al., 2003; Suzuki et al., 2006; Zaphiropoulos, 1997). However, the advent of next generation sequencing has illuminated circRNAs as an entire class of abundant, non-coding RNAs ubiquitous among eukaryotes (Guo et al., 2014; Jeck and Sharpless, 2014; Jeck et al., 2013; Lasda and Parker, 2014; Memczak et al., 2013; Salzman et al., 2012, 2013; Wang et al., 2014; Wilusz and Sharp, 2013; Zhang et al., 2014). circRNAs result from a non-canonical form of alternative splicing, most commonly where the splice donor site of one exon is ligated to the splice acceptor site of an upstream exon (Figure 1). Lacking 3' termini, circRNAs are non-polyadenylated and resistant to

digestion of the RNA with RNase R, a highly processive 3' exonuclease that non-specifically degrades linear RNA, but not circRNAs, attributes which are exploited in their sequencing and identification (Jeck and Sharpless, 2014).

The finding that circRNAs are widespread in human and animal tissues raises two important questions: what controls their formation, and what are their function(s) (if any)? Because only two specific circRNAs, *cIRS-7/CDR1as* and *Sry*, have had any function ascribed to date, both acting as micro (mi)RNA sponges (Hansen et al., 2013; Memczak et al., 2013), it remains possible that the majority of circRNAs are accidental by-products of mis-splicing. However, mining of ENCODE sequence data has revealed that patterns of circRNA expression can be cell-type specific, suggesting their formation may be regulated, which in turn would indicate they have functions (Salzman et al., 2013). Recent reports indicate circRNA formation can be aided by the close proximity of circRNA splice sites mediated by complementary base pairing of inverted repeats in the introns flanking the circRNA-forming exons and that many circRNAs rely on this for their biogenesis (Ashwal-Fluss et al., 2014; Liang and Wilusz, 2014; Zhang et al., 2014). However, this intron-pairing phenomenon alone cannot explain how a single abundant transcript common to a multitude of cells can host cell-type specific circRNAs (Jeck and Sharpless, 2014; Salzman et al., 2013). On the other hand, regulated alternative splicing plays a major role in expanding the transcriptome is critical in development and in physiological responses (Kalsotra and Cooper, 2011), making it likely that splicing factors may participate in regulating circRNA biogenesis. Indeed, production of a single circRNA from the pre-mRNA of the Muscleblind splicing factor was recently shown to be regulated by Muscleblind itself (Ashwal-Fluss et al., 2014). However, it is not known whether Muscleblind or other factors can regulate circRNAs on a wider scale.

Quaking, which belongs to the STAR family of KH domain-containing RNA binding proteins, has been found to affect pre-mRNA splicing (Hall et al., 2013; Wu et al., 2002), mRNA turnover (Larocque et al., 2005), and translation (Saccomanno et al., 1999) and has been implicated in diseases including ataxia, schizophrenia, and glioblastoma (Chénard and Stephane, 2008). The Quaking (QKI) gene is processed into three major isoforms of 5, 6, and 7 Kb, called QKI-5, QKI-6, and QKI-7, respectively, that have substantially different 3'UTRs, but differ by only about 30 amino acids in their C-termini and all of which retain the KH





### Figure 1. circRNAs Are Regulated in Human EMT

(A) Schematic depiction of alternative splicing isoforms generated from linear splicing (top) and back-splicing (bottom) of a four-exon transcript. The typical locations of divergent primers used for quantitation of circRNAs by RT-PCR are shown as blue arrows.

(B) Phase contrast images of HMLE (epithelial) cells and mesHMLE (mesenchymal) cells resulting from TGF- $\beta$  treatment for 21 days to induce EMT. Scale bar, 50  $\mu$ m. Shown below is the relative abundance of prototypical EMT-regulated transcripts in HMLE and mesHMLE cells as determined by qRT-PCR normalized to GAPDH. Values are mean  $\pm$  SEM, n = 3.

(C) Distributions of total circRNAs (left) and circRNAs arising from abundant transcripts (FPKM >1.0) that were changed by <25% in EMT (right).

(D) Correlation of fold-change in abundance of each circRNA (y axis) and its cognate mRNA (x axis) following EMT. The individual distribution profiles of each are shown above and to the right of the correlation plot.

(E) Relative abundance of circRNAs in HMLE and mesHMLE cells as determined by qRT-PCR normalized to GAPDH. Values are mean  $\pm$  SEM, n = 3. See also [Figures S1](#) and [S2](#); [Tables S1](#) and [S2](#).

RNA binding domain. QKI 5, the most abundant isoform, is predominantly nuclear, while QKI-6 can be nuclear and cytoplasmic, and QKI-7 is predominantly cytoplasmic (Pilotte et al., 2001). QKI dimerizes through its N-terminal Qua1 domain (Teplova et al., 2013) and binds bipartite sequence motifs (Galarneau and Richard, 2005) that can be on the same or separate RNA molecules (Teplova et al., 2013). PAR-CLIP crosslinking analysis in human embryonic kidney cells (HEK293T) indicates the majority of QKI binding occurs within introns, consistent with a role in splicing (Hafner et al., 2010).

Epithelial-mesenchymal transition (EMT) is a cellular differentiation process important in embryo development, wound healing, and in cancer metastasis (Nieto, 2013). The differentiation of an epithelial cell into a mesenchymal cell involves drastic changes in cell morphology, in gene expression patterns, and in arrange-

ment and function of the actin cytoskeleton (Bracken et al., 2014). EMT can be triggered by various ligand-receptor interactions, including TGF- $\beta$ , Wnt, and FGF, and involves extensive regulatory networks that are controlled by transcription factors and miRNAs (Lamouille et al., 2014). Because many cancers arise from epithelial cells and need to undergo EMT to become invasive and to metastasise, understanding the regulatory processes involved in EMT may reveal new modalities for therapeutic intervention in cancer progression.

We report here that the expression of hundreds of circRNAs is regulated during EMT in response to TGF- $\beta$ , with the majority of regulated circRNAs increasing in abundance. To screen for RNA binding proteins that regulate circRNA formation, we devised a dual color reporter construct, called *circScreen*, allowing simultaneous quantification of linear and circRNA splicing. Using

*circScreen*, we identified the RNA binding protein Quaking (QKI) as a major regulator of circRNA biogenesis in EMT. Furthermore, we show that introduction of consensus binding sequences for QKI into the flanking introns is sufficient to cause circRNAs to be produced from exons that normally only undergo canonical linear splicing. Because some of the most highly expressed circRNAs are among those that are regulated in EMT, our findings strongly suggest that certain circRNAs have EMT-related functions and thus may affect mesenchymal cell properties such as migration, invasion, and the propensity for cancers to metastasise.

## RESULTS

### circRNA Formation in EMT

To assess whether circRNA production is regulated in EMT, we harvested RNA from immortalized human mammary epithelial (HMLE) cells before and after they had undergone EMT in response to treatment with TGF- $\beta$ . The TGF- $\beta$ -treated cells are stably mesenchymal with typical morphology and marker expression and are referred to as mesHMLE cells (Figure 1B) (Attema et al., 2013; Mani et al., 2008). RNA from two biological replicates of HMLE and mesHMLE cells was subjected to deep sequencing, using library preparation procedures and bioinformatics pipelines designed to detect circRNAs (refer to [Experimental Procedures](#)). We detected 5,178 distinct circRNAs in HMLE and 8,084 circRNAs in mesHMLE cells, of which 3,417 were common to both cell types (Figure 1C; Table S1). These circRNAs are produced from 3,632 genes, meaning that circRNAs are not pervasive, rather being produced from approximately 22.5% of the transcriptome (16,138 genes with FPKM >0.1) in these cells. To validate the circRNA analysis pipeline, we selected a subset of circRNAs that vary in their abundance, size, and genetic location and performed RT-PCR on RNase R-treated RNA from HMLE and mesHMLE cells using circRNA-specific divergent primers (Figure 1A). We successfully amplified the expected product from 75/78 predicted circRNAs (Figure S1; Tables S1 and S2), authenticating our circRNA prediction pipeline.

To check that the increase in circRNAs in mesHMLE cells was a consequence of active regulation rather than a consequence of increased transcription of the parent gene, we quantified circRNAs arising from abundant transcripts (FPKM >1.0) that were largely unchanged (< 25% change) in EMT. The relative proportions of these circRNAs in HMLE and mesHMLE cells were similar to circRNAs derived from all transcripts (Figure 1C), indicating that substantially more of these circRNAs were present in mesenchymal cells than in epithelial cells. To assess whether there was a tendency for circRNAs that are present in HMLE cells to increase in abundance following EMT, we analyzed all abundant circRNAs common to HMLE and mesHMLE cells and plotted the fold change in circRNA abundance against the fold change in abundance of the cognate mRNA (Figure 1D). While the mRNA fold changes exhibited a symmetrical, normal distribution in response to TGF- $\beta$  treatment, the circRNA profile was positively shifted toward increased abundance (Figure 1D). To verify this regulation, we performed quantitative (q)RT-PCR using primers designed to measure several of the strongly regu-

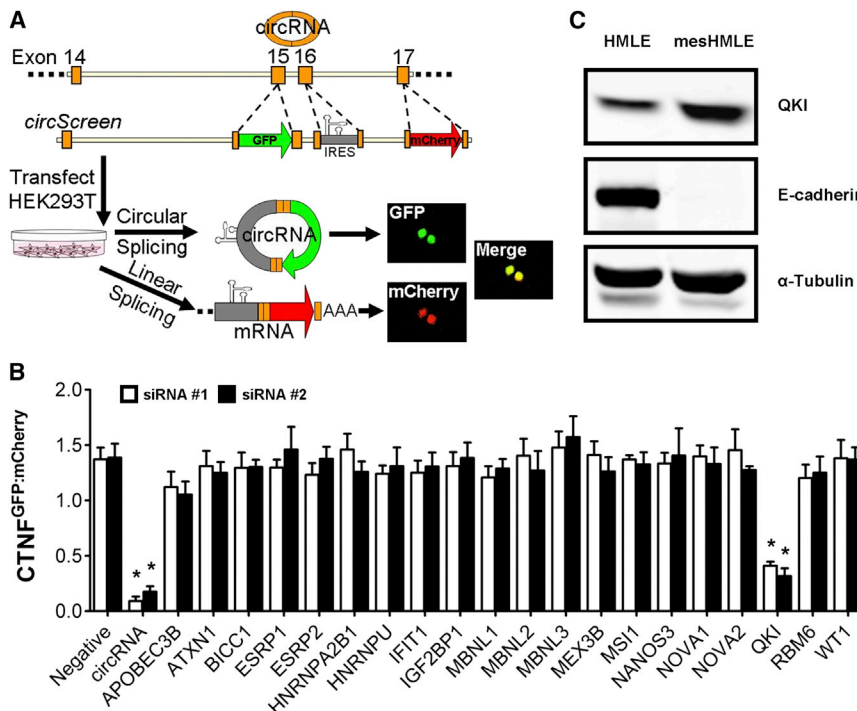
lated circRNAs, which confirmed that the circRNAs from POLE2, OXNAD1, SHPRH, SMAD2, and ATXN2 were all increased over 40-fold by the TGF- $\beta$  treatment, while circRNAs from DOCK1 and GNB1 were strongly decreased (Figure 1E).

To assess whether circRNA formation might be a mechanism to encourage long range exon skipping in the parent transcript, we analyzed the RNA sequencing data for evidence of mRNA isoforms of SMARCA5, POLE2, SHPRH, SMAD2, and ATXN2 in which the circRNA-forming exons are skipped. We found in most samples there were no reads corresponding to this form of exon skipping. Where such exon skipping was detected, it corresponded to <1% of the non-skipped mRNA isoform. Thus, regulated circRNA formation does not appear to be a mechanism for regulating exon skipping.

To check that the increases in circRNAs were not due to a reduced cell proliferation rate following EMT, allowing more accumulation of stable circRNAs, we measured the proliferation rates of the HMLE and mesHMLE cells and found the proliferation rate of mesHMLE cells is actually higher than that of HMLE cells (Figure S2). Thus, the increases in circRNA levels are due to increased biogenesis. Furthermore, given that >62% of circRNAs were cell-type-specific, being detected in only HMLE or mesHMLE cells (Figure 1C), despite the similar abundance of their cognate mRNA, this precludes complementary base-pairing in introns alone as the mechanism of biogenesis for the majority of circRNAs in these cells. Rather, these results suggest substantial regulation of circRNA biogenesis, independent of changes in abundance of the parental transcript during EMT.

### A Focused Screen for Regulators of circRNA Formation

The regulation of circRNA formation during EMT suggests there may be regulatory factors that participate in circRNA biogenesis. We postulated that if a protein factor contributes to circRNA biogenesis, it would also be regulated over the same time-course. To provide a tool to screen for regulatory factors that affect circRNA formation, we constructed a dual color fluorescent reporter, called *circScreen*, enabling simultaneous quantification of both linear and circRNA splicing from a minigene reporter construct (Figure 2A). The *circScreen* reporter construct incorporates exons 14–17 of the SMARCA5 gene, a member of the SWI/SNF family of chromatin remodelling proteins, which are as frequently mutated in cancer as p53 (Wang et al., 2014b), primarily chosen because (1) this region produces an abundant circRNA comprised of exons 15 and 16, the abundance of which was increased 8-fold following EMT (Figure 1E) with minimal change in the abundance of the linear mRNA transcript (FPKM<sup>HMLE</sup> = 23.0, FPKM<sup>mesHMLE</sup> = 26.9), (2) it lacks complementary *Alu* elements, and (3) it is comprised of sufficiently small exons and introns to permit a modest-sized reporter plasmid. The reporter was constructed such that the normally untranslated circRNA produced from the minigene undergoes IRES-mediated translation of GFP, while the linear mRNA that is produced gives rise to red fluorescing mCherry (Figure 2A). Both fluorescent proteins are tagged (GFP-FLAG, mCherry-HA) and have nuclear localization signals inserted, confining them to the nucleus to aid quantitation of cellular fluorescence (analysis pipeline summarized in Figure S3). We verified that the reporter was accurately spliced (Figure S4A) and gave rise to nuclear



**Figure 2. Identification of QKI as a circRNA Biogenesis Factor Using the *circScreen* Reporter**

(A) Schematic of the *circScreen* reporter construct and fluorescence images of cells transfected with *circScreen*. Nuclear green fluorescence is a marker of IRES-driven translation from circRNA; nuclear red fluorescence results from IRES-driven translation from the linear mRNA.

(B) Effect of siRNA-mediated knockdown of EMT-regulated RBPs on the ratio of circular to linear RNAs produced from the *circScreen* reporter, as measured by ratio of green to red corrected total nuclear fluorescence (CTNF). There were two different siRNAs that were used per target, shown in white and black bars. Data were acquired from measurements on 300 cells per experiment and are presented as mean  $\pm$  SEM,  $n = 3$ .

(C) Western blot of QKI, E-cadherin, and  $\alpha$ -tubulin (loading control) levels in HMLE and mesHMLE cells. See also Figures S3 and S4.

### QKI Regulates Formation of circRNAs via Binding Sites in Introns

The effect of QKI on circRNA formation could conceivably be indirect, or could

be through direct binding of QKI to the pre-mRNA. To assess whether QKI binds the SMARCA5 pre-mRNA, we performed RNA-immunoprecipitation (RIP) assays (Figure S5), using qRT-PCR to quantify QKI occupancy within the introns adjacent to the circRNA-forming exons. We found that QKI binds to the exon-adjacent sites at a level comparable to its binding to a site in the previously validated QKI target, NUMB (Zong et al., 2014), whereas binding to more remote regions elsewhere in SMARCA5 and to the circRNA itself was negligible (Figures 3A and 3B). To assess whether QKI binding sites in the introns flanking the circRNA-forming exons of SMARCA5 are necessary for circRNA biogenesis, we searched for sequences that match potential QKI response elements in the vicinity of the QKI RIP-enriched regions (Figures 3A and 3B) and found four instances of a bipartite motif that contains the sequence UAAY in conjunction with a relaxed version of the canonical QKI hexamer previously determined by SELEX (Galarneau and Richard, 2005). There are two of the putative elements that are located upstream and two are located downstream of the circRNA-forming splice sites, as shown in Figure 3C. Mutation of any of the putative binding sites individually had little effect on circRNA formation, but mutation of both members of either the upstream pair or the downstream pair substantially reduced circRNA formation, while mutation of all four sites was even more effective (Figure 3D). Knockdown of QKI strongly reduced circRNA formation from the intact or singly mutated reporter, but not from the reporter with both members of the pair mutated, confirming the absence of off-target effects on the reporter (Figure S4E). Together, these data indicate that QKI binds upstream and downstream of the circRNA-forming exons in SMARCA5 to promote circRNA formation.

green fluorescence that could be almost eliminated by co-transfection of either of two small interfering (si)RNAs targeting the exon16/exon15 junction unique to the circRNA, without significantly affecting mCherry levels (Figure 2B), as confirmed by qPCR and western blot (Figure S4B).

To identify splicing-associated factors involved in circRNA biogenesis, we first selected a candidate panel of nuclear RNA binding proteins (RBPs) (based on Gene Ontology annotations) that were expressed at appreciable levels ( $>0.1$  FPKM in polyA<sup>+</sup> RNA-sequencing from at least one sample) and were changed in abundance by  $>2$ -fold, either up or down, following EMT (Table 1). We also incorporated siRNAs that target the three human homologs of *Drosophila Muscleblind* (MBNL1, MBNL2, and MBNL3) implicated in circRNA biogenesis from its cognate locus (Ashwal-Fluss et al., 2014). We examined the effect of siRNA-mediated knockdown of each of these RBPs on circRNA formation using the *circScreen* minigene reporter in the readily transfectable HEK293T cell line, which has been shown to be capable of circRNA formation from the SMARCA5 locus (Memczak et al., 2013). The knockdown of most members of the RBP panel, and all three MBNL homologs, had little effect on the ratio of GFP:mCherry, but knockdown of QKI caused a substantial decrease in the ratio, indicating the QKI protein is required for efficient formation of circRNA from this reporter (Figure 2B). The decrease in circRNA production on knockdown of QKI was confirmed by western blotting and qRT-PCR, which showed a reduction in tagged GFP reporter protein and in circRNA, but not linear RNA from the transgenic and endogenous genes (Figures S4C and S4D). Importantly, we confirmed that QKI protein is increased in the cells that have undergone EMT (Figure 2C) and is knocked down by  $>90\%$  by siRNA-mediated silencing (Figure S4E).

**Table 1. EMT-Regulated Nuclear RBPs**

RBP Gene Name	Fold-Change (Log <sub>2</sub> )
ESRP1	−9.4
ESRP2	−2.3
ATXN1	1.1
QKI	1.1
WT1	2.1
BICC1	2.1
APOBEC3B	2.1
IFIT1	2.3
NANOS3	2.6
IGF2BP1	2.8
NOVA1	3.4
MSI1	3.5
MEX3B	3.7
NOVA2	7.1

To more broadly assess the role of QKI in circRNA production, we examined the effect of QKI knockdown on the abundance of 13 circRNAs that were previously identified in HEK293T cells (Memczak et al., 2013) and compared this with the presence or absence of known QKI PAR-CLIP sites (Hafner et al., 2010) in the introns flanking the circRNA-forming exons. The abundance of all nine circRNAs with adjacent QKI PAR-CLIP sites was reduced following QKI knockdown, whereas the four that are devoid of adjacent sites were largely unaffected (Figure 3E). To extend the identification of QKI-dependent circRNAs in the EMT context, we performed RNA sequencing on RNA from two biological replicates of mesHMLE cells in which QKI was knocked down by siRNA treatment and the abundances of the circRNAs were compared with those in control mesHMLE cells. We found that there was a strong bias toward decrease in circRNA abundance (Figure 3F). Of the 300 most abundant circRNAs in mesHMLE cells, 105 were decreased more than 2-fold by QKI knockdown, whereas only seven were increased by more than 2-fold (Figure 3F). We also confirmed by qRT-PCR that the increase in circRNAs from SMARCA5, POLE2, OXNAD1, SHPRH, SMAD2, and ATXN2 that occurred in EMT (Figure 1E) was abrogated by QKI knockdown, while DOCK1 and GNB1, which lack QKI response elements (QREs) in the adjacent introns, were unaffected by QKI knockdown (Figure S4I). To investigate which of the three isoforms of QKI is responsible for circRNA formation, we knocked down each isoform individually in mesHMLE cells and measured the levels of the QKI-dependent SMARCA5, POLE2, SHPRH, SMAD2, and ATXN2 circRNAs by qRT-PCR. The isoform-specific siRNAs that target QKI-6 and QKI-7 had little effect on these circRNA levels, whereas the QKI-5-specific siRNA reduced the level of all five circRNAs (Figures S4F and S4G), indicating it is QKI-5 that is responsible for the circRNA formation, consistent with this nuclear isoform acting on circRNA formation during splicing. Consistent with a role for QKI binding in regulating circRNA production in EMT, we found, using the PAR-CLIP data of Hafner et al. (2010) to indicate potential QKI binding sites, that the EMT-regulated circRNAs were

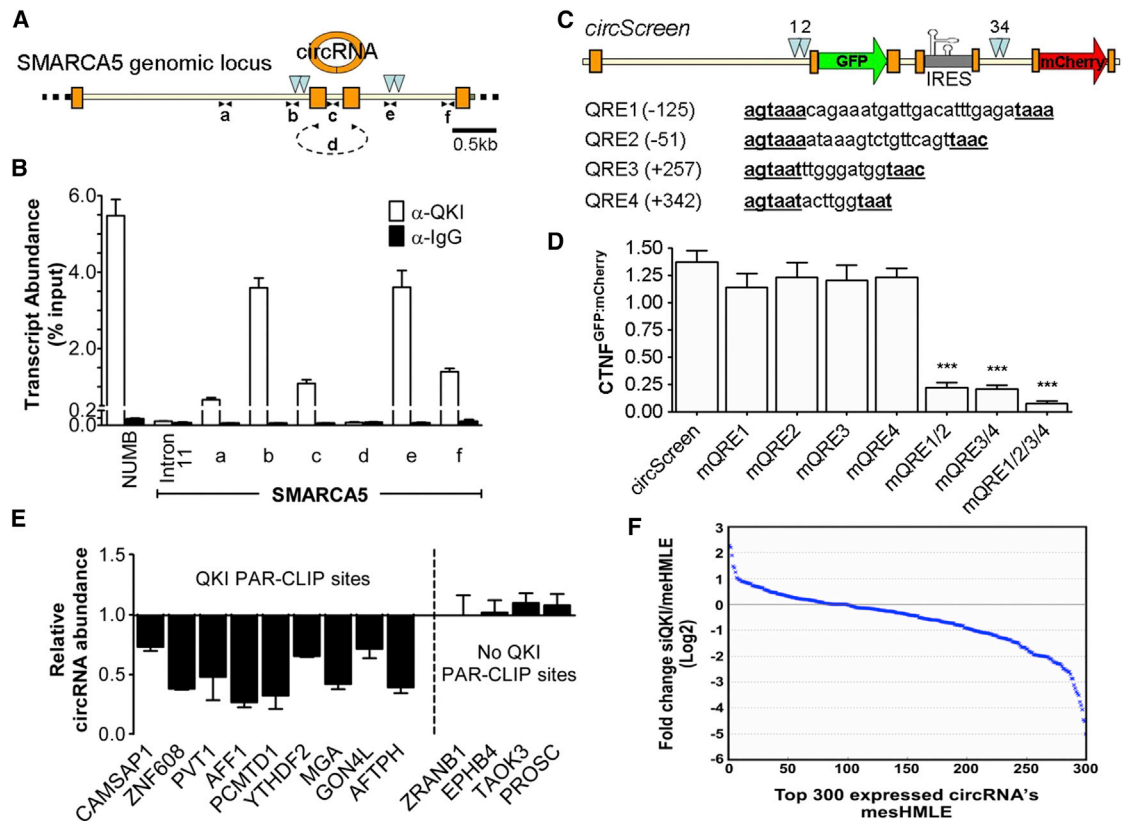
1.3-fold more likely than unregulated circRNAs to have a QKI site in a flanking intron and 2.1 times more likely to have QKI sites in both flanking introns. Together, these data demonstrate that QKI plays a major role in regulating circRNA production during EMT.

To establish that circRNA regulation by QKI is sustained for longer periods, QKI was stably knocked down by two different QKI-specific small hairpin (sh)RNAs in mesHMLE cells, reducing QKI protein by >85% (Figure 4A). Compared to a scrambled universal shRNA, the change in abundance of circRNAs measured by qRT-PCR mimicked that for the siRNA (Figure 4B). Interestingly, overexpression of QKI in mesHMLE cells (Figure 4C) increased the circRNA abundance in two independent clones (Figure 4D), revealing a quantitative positive correlation between QKI abundance and circRNA biogenesis. This was confirmed beyond the HMLE EMT model by performing qRT-PCR on four circRNAs (SMARCA5, POLE2, SHPRH, and DOCK1) in the prototypical epithelial (MCF7) and mesenchymal (MDA-MB-231) breast cancer cell lines. The circRNAs from SMARCA5, POLE2, and SHPRH, which are flanked by QKI PAR-CLIP motifs, were more abundant in the mesenchymal cell line, which express over 2.5-fold more QKI transcript than MCF7 cells, while DOCK1 was more abundant in the epithelial cell line (Figure S6A). This indicates that the relationship between QKI and circRNA production is maintained in breast cancer cells.

#### Insertion of Synthetic QKI Binding Sites into Introns Is Sufficient to Generate circRNA Formation

As a conclusive test of QKI-directed circRNA biogenesis, we investigated whether exons that do not normally produce circles could be made competent to produce circRNA by insertion of QKI binding motifs into the adjacent introns. We selected four genes, SYT8, ADD3, TIMP1, and NACAD, that are expressed, but do not give rise to circRNAs in HMLE, mesHMLE, MDA-MB-231, or HEK293T cells. Minigene expression vectors were constructed to express a region encompassing three exons from each of these genes, with and without canonical QKI binding motifs (ACUAAACN<sub>1–20</sub>UAAC motif determined by SELEX, see Galarneau and Richard, 2005) inserted in both introns flanking the central exon (Figure 5A). The minigenes were expressed in HEK293T cells (Figure 5) and MDA-MB-231 cells (Figures S6B and S6C) and assayed for circRNA formation by RT-PCR using divergent primers that can only give a product from circularized RNA (Figure 5B). None of the unmodified minigenes was capable of producing a circRNA, including ADD3, which contains two QKI PAR-CLIP sites 5' to the central exon (Hafner et al., 2010). However, insertion of another intronic QRE downstream of the central exon in ADD3, and insertion of QREs into both flanking introns for the other three reporters resulted in production of circRNA from each minigene (Figure 5B), verified by sequencing of the RT-PCR products and by their resistance to treatment with RNase R, an exonuclease that degrades linear, but not circular, RNA (Figure 5B). To verify that the circRNA production was dependent on QKI, the effect of siRNA-mediated QKI knockdown was tested, and it was found that this largely abrogated circular, but not linear product formation in the RT-PCR (Figure 5B), consistent with circRNA production being dependent on QKI. Together, our data show that QKI promotes circRNA production





**Figure 3. QKI Binds to Pre-mRNA to Stimulate circRNA Biogenesis**

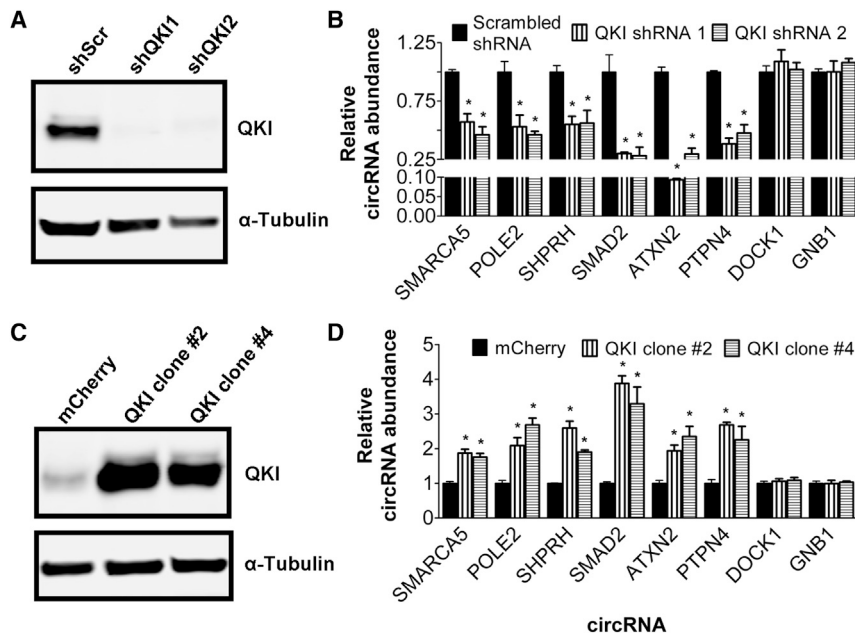
(A) Schematic of SMARCA5 pre-mRNA showing the locations of four putative QREs (inverted blue triangles) and amplicons (A–F) used for RIP assay. (B) RIP assay using the PCR primers indicated in (A). The validated QKI binding site in *NUMB* was used as a positive control. (C) Schematic diagram showing locations and sequence of predicted QREs in the *circScreen* reporter gene. Numbers in brackets refer to the distance from the circRNA forming splice site. (D) Effect of mutations to the QREs on the ratio of circRNA to linear mRNA, as determined by ratios of CTN<sup>GFP:mCherry</sup>. \*\*\*p < 0.001. Data were acquired from measurements on 300 cells per experiment and are presented as mean ± SEM, n = 3. (E) Effect of QKI knockdown on abundance of various circRNAs in HEK293T cells. The presence or absence of QKI binding sites in these genes, as determined by PAR-CLIP assay by Hafner et al. (2010) is indicated. circRNAs were measured by qRT-PCR with data presented as circRNA abundance relative to negative siRNA control cells, mean ± SEM, n = 3. (F) Ranked fold changes in circRNA abundance for the 300 most highly expressed circRNAs in mesHMLE cells following siRNA-mediated knockdown of QKI. Data are means from two replicate experiments. See also Figures S4 and S5.

from genes that have QKI binding sites appropriately located within the introns.

## DISCUSSION

From deep sequencing of RNA, we have detected thousands of circRNAs present in HMLE epithelial cells and in their mesenchymal counterparts that are formed in response to prolonged exposure to TGF-β. We verified that the sequence analysis pipeline appropriately identifies circRNAs by performing RT-PCR on RNase R-treated RNA, using divergent primers that amplify across the splice junction unique to the circular form of the RNA. Consistent with previous studies on different cell types (Guo et al., 2014; Jeck et al., 2013; Salzman et al., 2013), we found that the majority of circRNAs in epithelial cells, and in their mesenchymal derivatives, are of low abundance and consequently could conceivably be the result of errors in pre-mRNA

splicing, but certain circRNAs were present at substantial levels that suggest they are purposefully produced. Furthermore, the production of numerous abundant circRNAs was regulated in response to TGF-β, independent of changes in their cognate mRNA transcript, which strongly suggests they are produced to carry out some function in these cells. Most of the circRNAs whose level changed substantially in EMT were upregulated, suggesting they carry out functions related to the mesenchymal phenotype. A smaller number were also strongly changed in the opposite direction, consistent with these circRNAs having epithelial-specific functions. For example, the *DOCK1* circRNA is one of the most abundant circRNAs in the epithelial cells, but was downregulated 30-fold in response to TGF-β. In light of the observation that circRNA production competes with linear splicing of pre-mRNA (Ashwal-Fluss et al., 2014), it is interesting to note that the level of *DOCK1* mRNA was increased (by about 2-fold) in response to TGF-β, raising the possibility that one



**Figure 4. QKI Perturbs Numerous circRNAs in mesHMLE**

(A) Western blot of QKI and  $\alpha$ -tubulin (loading control) levels in mesHMLE cells stably transduced with pLKO::scrambled shRNA (shScr), pLKO::QKI shRNA 1 (shQKI1), and pLKO::QKI shRNA 2 (shQKI2).

(B) qRT-PCR of circRNAs from mesHMLE cells with shRNA constructs. Data presented as circRNA abundance relative to scrambled shRNA control cells, mean  $\pm$  SEM,  $n = 3$ .

(C) Western blot of QKI and  $\alpha$ -tubulin (loading control) levels in mesHMLE cells stably transduced with pLX301::mCherry, pLX301::QKI clone #2, and pLX301::QKI clone #4.

(D) qRT-PCR of circRNAs from mesHMLE cells with QKI overexpression. Data presented as circRNA abundance relative to pLX301::mCherry control cells, mean  $\pm$  SEM,  $n = 3$ .

function of excision of the DOCK1 circRNA could be to contribute to downregulation of DOCK1 mRNA in epithelial cells. DOCK1 is a guanine nucleotide exchange factor (GEF) that activates Rac to enhance cell motility (Gadea and Blangy, 2014), consistent with its upregulation in EMT. Nevertheless, because the DOCK1 circRNA is so abundant in epithelial cells, it is tempting to speculate that it has a function in these cells apart from a role in reducing expression of the DOCK1 mRNA.

Our finding of frequent regulation of circRNA abundance in EMT is a strong argument in favor of these circRNAs having functions in the cell, although these functions remain largely unknown. Just two circRNAs have had functions ascribed to them to date: ciRS-7/CDR1as acts as a sponge for miR-7 in mammalian cells, and a circRNA from the testes-specific *Sry* gene acts as a sponge for miR-138 (Capel et al., 1993; Hansen et al., 2013; Memczak et al., 2013). We searched for reiterated miRNA binding sites in the circRNAs that are regulated by TGF- $\beta$  treatment of HMLE cells, but did not find any notable examples, suggesting they do not act as miR sponges. This observation is in agreement with the report of Guo et al. (2014) that the majority of circRNAs appear to not function as miRNA sponges.

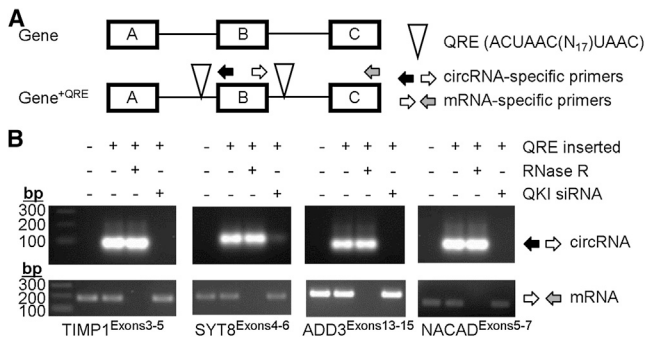
We observed that changes in circRNA levels were mostly in the direction of increased abundance in the mesenchymal cells, but at least for the more abundant circRNAs, their expression was not exclusively mesenchymal. Several abundant circRNAs were identified to change by 4–10-fold, while about 50 changed by 2–4-fold, with many of these having substantial abundance in the epithelial cells (Table S1). This contrasts with the almost exclusively mesenchymal expression of key transcription factors (such as ZEB1, Twist1, and Snail) that drive the mesenchymal gene expression program and suggests that the functions of the upregulated circRNAs are not likely to be exclusive to mesenchymal cells. More relevant comparisons in this regard may be to genes that are expressed in epithelial cells, but are

increased in mesenchymal cells because they participate in activities such as controlling cell shape, ECM interactions, or migration, which are more prominent in

the mesenchymal phenotype. For example, some tubulin, cofilin, and laminin transcripts have these features and were moderately increased in EMT.

Because circRNAs are likely to be quite stable, it is tempting to speculate that the functions of some circRNAs may take advantage of their long half-lives, allowing them to act as slow-responding regulators. A long half-life confers a slow approach to steady-state level and a slow decline if their production ceases. Furthermore, if the degradation rate is considerably longer than the rate of cell division, the steady-state level becomes sensitive to the cell division rate, which effectively replaces degradation in determining the abundance of the circRNA. EMT is a rather special type of induced cellular response in that it causes a profound change in the differentiation state of the cell, but at least in vitro, is readily reversible in the initial days of induction, and then becomes more refractory to reversion, but nevertheless remains reversible (Gregory et al., 2011). This has parallels in vivo as exemplified by the reversals of EMT that occur in some tissues during embryological development, while reversible plasticity is thought to play an important role in metastatic colonization (Brabletz, 2012; Nieto, 2013). Perhaps some circRNAs that are induced during EMT could be involved in helping to determine the slow development, or the duration, of a refractory mesenchymal state.

We have found that a substantial contribution to the regulation of circRNA production in EMT comes from the regulation of circularization by the mesenchymal splicing factor, QKI. QKI is essential for enhanced production of many circRNAs and acts by binding to recognition elements within introns, in the vicinity of the circRNA-forming splice sites. Furthermore, insertion of QKI motifs is sufficient to induce circRNA formation in contexts where this is normally not observed. Secondary structure within pre-mRNAs that brings circRNA-forming exons into close proximity has been shown to enhance circRNA biogenesis (Liang and



**Figure 5. Introduction of QKI-Binding Sites Promotes Novel circRNA Formation**

(A) Schematic showing sites of insertion of QREs and locations of PCR primers used for segments of four genes (SYT8, ADD3, TIMP1, and NACAD) devoid of circRNAs that were cloned into pcDNA3.1.

(B) Gel electrophoresis of RT-PCR products using circRNA-specific (upper panels) and mRNA-specific primers (lower panels) on RNA from HEK293T cells transfected with the indicated minigenes, with siRNA-mediated silencing of QKI or treatment with RNase R. See also Figure S6.

Wilusz, 2014; Zhang et al., 2014). Since QKI is a dimer, capable of binding two well separated regions of a single RNA molecule (Teplova et al., 2013), it is an attractive possibility that QKI likewise promotes circRNA biogenesis by bringing the circle-forming exons into close proximity.

Our observations that QKI levels are regulated in the EMT produced by TGF- $\beta$  treatment of HMLE cells, and that QKI-regulated circRNAs similarly change in abundance, suggest that circRNAs could have important functions in EMT. Since EMT is widely regarded to have an important role in the progression of carcinomas to metastasis (Scheel and Weinberg, 2012; Tsai and Yang, 2013) it will be interesting to examine the influence of QKI-mediated circRNAs in cancers.

## EXPERIMENTAL PROCEDURES

### Cell Lines and Cell Culture

Human HMLE cell line cells were cultured and induced to undergo EMT as per Mani et al. (2008) using 2 ng/ml<sup>-1</sup> TGF- $\beta$ 1 (Sigma-Aldrich). HEK293T, MCF-7, and MDA-MB-231 immortalized breast cancer cells were maintained in Dulbecco's modified Eagle's medium (DMEM) GlutaMax (Life Technologies) with 10% (volume per volume) heat-inactivated fetal calf serum (Bovogen) and 1  $\times$  Penicillin-Streptomycin (Gibco) at 37°C with 5% CO<sub>2</sub>.

### qRT-PCR

Reverse transcription for mRNA and circRNAs were performed with SuperScript III First-Strand Synthesis System (Life Technologies) and 50 ng random hexamers or 2.5  $\mu$ M oligo (dT)<sub>20</sub> as per manufacturers instructions. miRNA qRT-PCR was performed with TaqMan assay as per Gregory et al. (2008). RT-PCR was subsequently performed in triplicate with a 1:10 dilution of cDNA with 1  $\times$  iQ SYBR Supermix (QIAGEN) on a Rotorgene 6000 series PCR machine (Corbett Research). Analysis was performed as per Gregory et al. (2008).

### Library Preparation

RNA was extracted from cells with TRIzol reagent (Life Technologies) and 4  $\mu$ g RNA was fractionated across three NEBNext Poly(A) mRNA Magnetic Isolation Module columns. The supernatant was retained as the polyA-depleted fraction (polyA<sup>-</sup>) and the bead-bound (polyA<sup>+</sup>) fraction eluted from beads in elution buffer/water according to manufacturer's instructions. The polyA<sup>-</sup> fraction was size-fractionated and concentrated using 3.5  $\times$  AMPure RNAClean XP sam-

ple preparation (New England Biolabs) and eluted with water. Ribo-Zero Magnetic Gold Kit (Human/Mouse/Rat) (Epicenter) treatment was performed on the polyA<sup>-</sup> fraction, purified, and size-fractionated with 1.7  $\times$  AMPure RNAClean XP beads. Stranded RNA libraries were made using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs) and multiplexed between 2- and 4-index on HiSeq 2500, 100 base pairs (bp) paired-end reads.

### Bioinformatics

Raw reads were adaptor trimmed and filtered for short sequences using cutadapt v1.3 (Martin, 2011), using minimum-length 18, error-rate 0.2, overlap five, paired-output options. The resulting FASTQ files were analyzed and quality checked using the FastQC program (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). The remaining reads were mapped against the GRCh37/hg19 human reference genome using the MapSplice spliced alignment algorithm (Wang et al., 2010) (version 2.1.8 beta, using parameters -bam -fusion -fusion-non-canonical -filtering 1 -min\_fusion\_distance 200-gene-gtf Ensembl). The number of reads supporting each circRNA junction was obtained from the resulting splice junction files.

Normalization was performed using a set of 165 linear non-polyadenylated "housekeeping" transcripts judged to be ubiquitously and consistently expressed across samples. This set included 39 small nucleolar RNA, 32 processed pseudogenes, 18 sense intronic elements, 18 snRNAs, eight large intergenic non-coding RNAs, and six miRNAs. Specifically, after taking logarithms, additive normalization factors were determined such that after normalization the sample-averaged expression of the housekeeping RNAs was identical to the overall expression average. These sample-specific normalization factors were then used to calculate appropriately normalized circRNA counts for each sample. As a cross-check of the normalization procedure we used, for a subset of the samples, a spike-in of a synthetic circRNA (refer to Extended Experimental Procedures for in vitro transcription of circRNAs). We found that, as expected, the normalization procedure tended to result in equilibrated spike-in levels.

### circScreen Reporter Construct

The genomic region of SMARCA5 was directionally cloned into pcDNA3.1 from HEK293T genomic DNA using primers SMARCA5\_BamHI\_F and SMARCA5\_NotI\_R (all DNA oligonucleotides in Table S1) with Phusion Hot Start Polymerase (Thermo Fisher Scientific) according to manufacturer's instructions. Cytomegalovirus internal ribosomal entry site (CMV-IRES) was amplified from pLMP-Cherry and cloned into the BstBI site of SMARCA5 Exon 16. GFP was amplified from pDendra2 vector (Clontech) to include N-terminal nuclear localization signal (NLS and KKRRKV) and C-terminal FLAG tag (DYKDDDK) and cloned into the BsaBI site of SMARCA5 Exon 15. NLS-mCherry-HA was generated from two overlapping gBlocks gene fragments (Integrated DNA Technologies) into PflFI sites, inserting mCherry into Exon 17 to complete the circScreen reporter. Mutation of QREs and deletion of inverted repeats from the circScreen reporter were achieved with DpnI-based (New England Biolabs) site-directed mutagenesis with Phusion DNA Polymerase. Sanger sequencing and restriction digestions confirmed sequence/frame and orientation of fragments.

### Nuclear RBP siRNA Screen

Cells were transfected with siRNAs at 5–20 nanomolar (nM) final concentration using Lipofectamine RNAiMAX (Life Technologies) in 24-well plates, while expression plasmids (1  $\mu$ g) were transfected with Lipofectamine 2000 (Life Technologies). For the circScreen reporter assay, siRNAs were transfected into HEK293T cells 40 hr prior to transfection with the circScreen plasmid. At five hours after circScreen transfection, the culture media was changed to 1:3 DMEM:F-12 (Life Technologies) with 10% heat-inactivated fetal calf serum and 1  $\times$  Penicillin-Streptomycin to reduce green autofluorescence. Cells were imaged on the IncuCyte Zoom (Essen Bioscience) capturing phase, green, and red fluorescence images at 10 $\times$  magnification over the next 48 hr (refer to Figure S3 for detailed image analysis).

### Incorporation of QREs into Minigene Reporters

There was four genes—SYT8, TIMP1, NACAD, and ADD3—that were chosen, as they were devoid of any circRNAs. Genomic regions comprising three

exons and two introns were synthesized (Integrated DNA Technologies) with and without consensus QREs (ACUAAAC(N<sub>17</sub>)UAAC) approximately 100–200 bp from the splicing site of the central exon. These were cloned into pcDNA3.1 and transfected into HEK293T cells that had been transfected 40 hr previous with negative control siRNA or QKI ON-Targetplus SMART pool siRNA (Dharmacon). RNA was isolated 24 hr later with TRIzol and treated with Ribonuclease R (RNase R, Epicenter) as per Jeck et al. (2013). cDNA was reverse transcribed using random hexamers and SuperScript III reverse transcriptase (Life Technologies).

### RNA Immunoprecipitation

MDA-MB-231 cells were UV cross-linked at 600 mJ/cm<sup>−2</sup> and lysed in 1× PXL lysis buffer (1× PBS, 0.1% SDS, 0.5% sodium deoxycholate, and 0.5% NP-40 with 1× Protease Inhibitor Cocktail [Promega]). Lysate was treated with DNase I (Roche) at 37°C for 10 min and centrifuged at 12,000 g for 30 min. Total protein (1 mg) was indirectly immunoprecipitated with 5 µg of α-QKI-5 rabbit polyclonal antibody (Bethyl Laboratories, A300-183A), or rabbit anti-mouse IgG antibody (Jackson Laboratories) as a control, and 100 µl of Protein G magnetic dynabeads (Life Technologies). Beads were washed twice with 1× PXL lysis buffer, followed by two washes each with 5× PXL and 1× PNK buffer (50 mM Tris-Cl [pH7.5], 10 mM MgCl<sub>2</sub>, and 0.5% NP-40). There was thirty percent of the immunoprecipitate that was set aside for western blot analysis, with protein eluted from beads in 1× PNK buffer, 4× LDS Loading Buffer (Life Technologies) and 4% β-mercaptoethanol at 70°C for 10 min. There was twenty percent of the immunoprecipitate that was treated for 20 min at 37°C with a Proteinase K solution (4 mg/ml Proteinase K in 100 mM Tris-Cl [pH 7.5], 50 mM NaCl, and 10 mM EDTA). An equal volume of Proteinase K solution including 7 M urea was then added and incubated at 37°C for a further 20 min, after which the RNA was extracted using phenol:chloroform. Samples were spun at 12,000 g, at 4°C for 10 min. The aqueous phase was collected and RNA precipitated overnight with 3 M sodium acetate [pH5.2], 20 µg glycogen, and 1:1 ethanol:isopropanol. The RNA pellet was washed twice with 75% ethanol, air-dried, and resuspended in water.

### Establishment of Cell Lines Stably Overexpressing QKI cDNA

Human QKI-5 cDNA was PCR amplified from pENTR-QKI (Addgene plasmid #16183) to include a stop codon and cloned into the BamHI and NotI sites of the pENTR2B entry vector (Invitrogen). The resulting pENTR2B-QKI-5 vector was recombined with pLX301 (Addgene plasmid #25895) using LR Clonase to generate pLX301-QKI-5. An expression vector containing mCherry (pLX301-mCherry) was generated by recombination of pENTR2B-mCherry with pLX301. Lentiviral particles were produced from pLX301-mCherry and QKI-5 constructs by co-transfecting HEK293T cells with pCMV5-VSVG and pCMV-dR8.2. A 1:8 dilution of virus was added to meshMLE cells plated at low density for 96 hr before selection with 1 µg/ml puromycin. Single clones that overexpressed QKI-5 and a control mCherry clonal pool were used for downstream experiments.

### Establishment of Cell Lines Stably Overexpressing QKI shRNA

Lentiviral vectors for shRNA mediated knockdown of QKI were purchased from Sigma-Aldrich (Mission pLKO; TRCN0000233372 and TRCN0000233375) along with a non-targeting control shRNA (Mission SHC216). Lentiviral particles and selection of stable meshMLE pools were carried out as described above for the pLX301 vectors.

### Western Blotting

Western blotting was performed according to standard protocols and imaged using the Odyssey CLx scanner (LI-COR). Primary antibodies used in this study were as follows: anti-α-tubulin mouse monoclonal (Abcam, ab7291, 1:10,000), anti-FLAG mouse monoclonal (Sigma, F1804, 1:4,000), anti-HA mouse monoclonal (Sigma, H3663, 1:1,000), anti-QKI5 rabbit polyclonal (Bethyl Laboratories, A300-183A, 1:5,000), and anti-E-cadherin mouse monoclonal (BD Biosciences, 610182, 1:1,000). Secondary antibodies used were, Goat anti-mouse IRDye680 (LI-COR, 926-32220, 1:20,000), Goat anti-rabbit IRDye680 (LI-COR, 926-32221, 1:20,000), Goat anti-mouse IRDye800 (LI-COR, 926-32210, 1:20,000); Goat anti-rabbit IRDye800 (LI-COR, 926-32211, 1:20,000).

### ACCESSION NUMBERS

The European Nucleotide Archive accession number for the RNA sequencing data reported in this paper is PRJEB8225.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.014>.

### AUTHOR CONTRIBUTIONS

S.J.C. designed the experiments with input from G.J.G. K.A.P., J.T., and A.W.S. performed the bioinformatic analyses. P.A.G. performed HMLE EMT time course and QKI knockdown experiments, purified RNA for qRT-PCR and library preparation, and provided important advice and reagents. V.C. assisted S.J.C. in synthesis of *circScreen*, performed *circScreen* quantification, and RT-PCR validation of targets. M.S. performed QKI RNA immunoprecipitation. C.A.P. generated QKI shRNA and overexpression lines with assistance from P.A.G. and S.R. S.R. performed western blotting, proliferation assays, and qRT-PCR across HMLE time course. All other experiments were performed by S.J.C. The figures were prepared by S.J.C., G.J.G., and J.T., and the manuscript was written by S.J.C. and G.J.G., with input from the other authors.

### ACKNOWLEDGMENTS

We thank all members of the Goodall Laboratory for valuable discussion. We thank David Lawrence of the CCB ACRF Cancer Genomics Facility for valuable discussion and for establishing sequencing analysis pipelines. Funding for this research was obtained from the Royal Adelaide Hospital Research Fund in the form of a Florey Postdoctoral Fellowship to S.J.C.; from the National Health and Medical Research Council project grants GNT1089167 to G.J.G. and S.J.C.; GNT1068773 to G.J.G. and P.A.G.; and from the EMPathy Breast Cancer Network, a National Breast Cancer Foundation (Australia)-funded National Collaborative Research Program. P.A.G. is supported by a Cancer Council SA Beat Cancer Fellowship.

Received: November 26, 2014

Revised: January 5, 2015

Accepted: February 3, 2015

Published: March 12, 2015

### REFERENCES

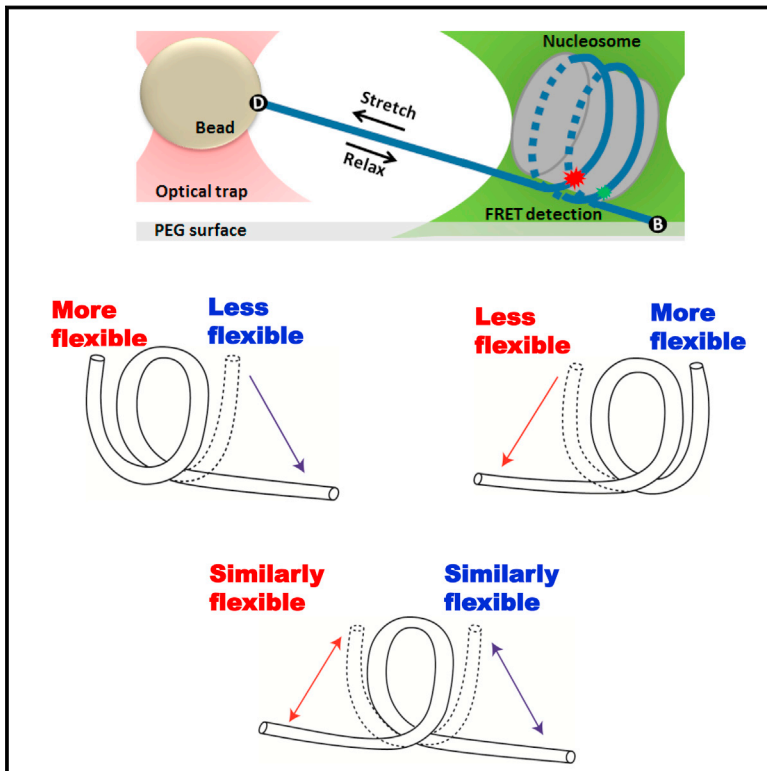
- Ashwal-Fluss, R., Meyer, M., Pamudurti, N.R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., and Kadener, S. (2014). *circRNA* biogenesis competes with pre-mRNA splicing. *Mol. Cell* 56, 55–66.
- Attema, J.L., Bert, A.G., Lim, Y.-Y., Kolesnikoff, N., Lawrence, D.M., Pillman, K.A., Smith, E., Drew, P.A., Khew-Goodall, Y., Shannon, F., and Goodall, G.J. (2013). Identification of an enhancer that increases miR-200b~200a~429 gene expression in breast cancer cells. *PLoS ONE* 8, e75517.
- Brabletz, T. (2012). To differentiate or not—routes towards metastasis. *Nat. Rev. Cancer* 12, 425–436.
- Bracken, C.P., Li, X., Wright, J.A., Lawrence, D.M., Pillman, K.A., Salamanidis, M., Anderson, M.A., Dredge, B.K., Gregory, P.A., Tsykin, A., et al. (2014). Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. *EMBO J.* 33, 2040–2056.
- Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P., and Lovell-Badge, R. (1993). Circular transcripts of the testis-determining gene *Sry* in adult mouse testis. *Cell* 73, 1019–1030.



- Chénard, C.A., and Stephane, R. (2008). New implications for the QUAKE RNA binding protein in human disease. *J. Neurosci. Res.* 86, 233–242.
- Cocquerelle, C., Mascres, B., Hétiu, D., and Bailleul, B. (1993). Mis-splicing yields circular RNA molecules. *FASEB J.* 7, 155–160.
- Gadea, G., and Blangy, A. (2014). Dock-family exchange factors in cell migration and disease. *Eur. J. Cell Biol.* 93, 466–477.
- Galarneau, A., and Richard, S. (2005). Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat. Struct. Mol. Biol.* 12, 691–698.
- Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* 10, 593–601.
- Gregory, P.A., Bracken, C.P., Smith, E., Bert, A.G., Wright, J.A., Roslan, S., Morris, M., Wyatt, L., Farshid, G., Lim, Y.Y., et al. (2011). An autocrine TGF-beta/miR-200 signaling network regulates establishment and maintenance of epithelial-mesenchymal transition. *Mol. Biol. Cell* 22, 1686–1698.
- Gualandi, F., Trabaneli, C., Rimessi, P., Calzolari, E., Toffolatti, L., Patarnello, T., Kunz, G., Muntoni, F., and Ferlini, A. (2003). Multiple exon skipping and RNA circularisation contribute to the severe phenotypic expression of exon 5 dystrophin deletion. *J. Med. Genet.* 40, e100.
- Guo, J.U., Agarwal, V., Guo, H., and Bartel, D.P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* 15, 409.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.-C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Hall, M.P., Nagel, R.J., Fagg, W.S., Shiue, L., Cline, M.S., Perriman, R.J., Donohue, J.P., and Ares, M., Jr. (2013). Quaking and PTB control overlapping splicing regulatory networks during muscle cell differentiation. *RNA* 19, 627–638.
- Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature* 495, 384–388.
- Jeck, W.R., and Sharpless, N.E. (2014). Detecting and characterizing circular RNAs. *Nat. Biotechnol.* 32, 453–461.
- Jeck, W.R., Sorrentino, J.A., Wang, K., Slevin, M.K., Burd, C.E., Liu, J., Marzluff, W.F., and Sharpless, N.E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* 19, 141–157.
- Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729.
- Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. *Nat. Rev. Mol. Cell Biol.* 15, 178–196.
- Larocque, D., Galarneau, A., Liu, H.-N., Scott, M., Almazan, G., and Richard, S. (2005). Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat. Neurosci.* 8, 27–33.
- Lasda, E., and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA* 20, 1829–1842.
- Liang, D., and Wilusz, J.E. (2014). Short intronic repeat sequences facilitate circular RNA production. *Genes Dev.* 28, 2233–2247.
- Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 133, 704–715.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12.
- Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., et al. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* 495, 333–338.
- Nieto, M.A. (2013). Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* 342, 1234850.
- Nigro, J.M., Cho, K.R., Fearon, E.R., Kern, S.E., Ruppert, J.M., Oliner, J.D., Kinzler, K.W., and Vogelstein, B. (1991). Scrambled exons. *Cell* 64, 607–613.
- Pilotte, J., Larocque, D., and Richard, S. (2001). Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKE apoptotic inducer. *Genes Dev.* 15, 845–858.
- Saccomanno, L., Loushin, C., Jan, E., Punkay, E., Artzt, K., and Goodwin, E.B. (1999). The STAR protein QKI-6 is a translational repressor. *Proc. Natl. Acad. Sci. USA* 96, 12605–12610.
- Salzman, J., Gawad, C., Wang, P.L., Lacayo, N., and Brown, P.O. (2012). Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* 7, e30733.
- Salzman, J., Chen, R.E., Olsen, M.N., Wang, P.L., and Brown, P.O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genet.* 9, e1003777.
- Scheel, C., and Weinberg, R.A. (2012). Cancer stem cells and epithelial-mesenchymal transition: concepts and molecular links. *Semin. Cancer Biol.* 22, 396–403.
- Suzuki, H., Zuo, Y., Wang, J., Zhang, M.Q., Malhotra, A., and Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* 34, e63.
- Teplova, M., Hafner, M., Teplov, D., Essig, K., Tuschl, T., and Patel, D.J. (2013). Structure-function studies of STAR family Quaking proteins bound to their in vivo RNA target sites. *Genes Dev.* 27, 928–940.
- Tsai, J.H., and Yang, J. (2013). Epithelial-mesenchymal plasticity in carcinoma metastasis. *Genes Dev.* 27, 2192–2206.
- Wang, K., Singh, D., Zeng, Z., Coleman, S.J., Huang, Y., Savich, G.L., He, X., Mieczkowski, P., Grimm, S.A., Perou, C.M., et al. (2010). MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* 38, e178.
- Wang, P.L., Bao, Y., Yee, M.-C., Barrett, S.P., Hogan, G.J., Olsen, M.N., Dinnyen, J.R., Brown, P.O., and Salzman, J. (2014a). Circular RNA is expressed across the eukaryotic tree of life. *PLoS ONE* 9, e90859.
- Wang, X., Haswell, J.R., and Roberts, C.W.M. (2014b). Molecular pathways: SWI/SNF (BAF) complexes are frequently mutated in cancer—mechanisms and potential therapeutic insights. *Clin. Cancer Res.* 20, 21–27.
- Wilusz, J.E., and Sharp, P.A. (2013). Molecular biology. A circuitous route to noncoding RNA. *Science* 340, 440–441.
- Wu, J.I., Reed, R.B., Grabowski, P.J., and Artzt, K. (2002). Function of quaking in myelination: regulation of alternative splicing. *Proc. Natl. Acad. Sci. USA* 99, 4233–4238.
- Zaphiropoulos, P.G. (1997). Exon skipping and circular RNA formation in transcripts of the human cytochrome P-450 2C18 gene in epidermis and of the rat androgen binding protein gene in testis. *Mol. Cell. Biol.* 17, 2985–2993.
- Zhang, X.-O., Wang, H.-B., Zhang, Y., Lu, X., Chen, L.-L., and Yang, L. (2014). Complementary sequence-mediated exon circularization. *Cell* 159, 134–147.
- Zong, F.-Y., Fu, X., Wei, W.-J., Luo, Y.-G., Heiner, M., Cao, L.-J., Fang, Z., Fang, R., Lu, D., Ji, H., and Hui, J. (2014). The RNA-binding protein QKI suppresses cancer-associated aberrant splicing. *PLoS Genet.* 10, e1004289.

# Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility

## Graphical Abstract



## Authors

Thuy T.M. Ngo, Qiucen Zhang, ...,  
Jaya G. Yodh, Taekjip Ha

## Correspondence

jyodh@illinois.edu (J.G.Y.),  
tjha@illinois.edu (T.H.)

## In Brief

Single-molecule fluorescence-force spectroscopy reveals that, despite its 2-fold symmetry in physical structure, the nucleosome unravels asymmetrically under tension and the direction of unwrapping is controlled by DNA flexibility.

## Highlights

- Nucleosome unwraps asymmetrically and directionally under tension
- Sequence-dependent DNA flexibility controls the unwrapping direction
- More flexible side is stronger mechanically and unwraps at higher forces
- Opening of one nucleosomal end helps to stabilize the other end



# Asymmetric Unwrapping of Nucleosomes under Tension Directed by DNA Local Flexibility

Thuy T.M. Ngo,<sup>1</sup> Qiucen Zhang,<sup>2</sup> Ruobo Zhou,<sup>2</sup> Jaya G. Yodh,<sup>2,\*</sup> and Taekjip Ha<sup>1,2,3,4,\*</sup>

<sup>1</sup>Center for Biophysics and Computational Biology

<sup>2</sup>Department of Physics, Center for Physics in Living Cells

<sup>3</sup>Carl R. Woese Institute for Genomic Biology

University of Illinois at Urbana-Champaign, Urbana, IL 61801-2902, USA

<sup>4</sup>Howard Hughes Medical Institute, University of Illinois, Urbana, IL 61801-2902, USA

\*Correspondence: jyodh@illinois.edu (J.G.Y.), tjha@illinois.edu (T.H.)

<http://dx.doi.org/10.1016/j.cell.2015.02.001>

## SUMMARY

Dynamics of the nucleosome and exposure of nucleosomal DNA play key roles in many nuclear processes, but local dynamics of the nucleosome and its modulation by DNA sequence are poorly understood. Using single-molecule assays, we observed that the nucleosome can unwrap asymmetrically and directionally under force. The relative DNA flexibility of the inner quarters of nucleosomal DNA controls the unwrapping direction such that the nucleosome unwraps from the stiffer side. If the DNA flexibility is similar on two sides, it stochastically unwraps from either side. The two ends of the nucleosome are orchestrated such that the opening of one end helps to stabilize the other end, providing a mechanism to amplify even small differences in flexibility to a large asymmetry in nucleosome stability. Our discovery of DNA flexibility as a critical factor for nucleosome dynamics and mechanical stability suggests a novel mechanism of gene regulation by DNA sequence and modifications.

## INTRODUCTION

The fundamental unit for genome compaction in eukaryotic cells is the nucleosome, in which ~147 base pairs of DNA wrap ~1.7 turns around a histone octamer core (Kornberg, 1974). Nucleosome dynamics regulates replication, repair, and transcription (Andrews and Luger, 2011; Bintu et al., 2012; Kulaeva et al., 2013; Li et al., 2007; Nag and Smerdon, 2009). Nucleosomal DNA can be invaded either passively due to spontaneous fluctuations (Hodges et al., 2009; Koopmans et al., 2007; Li et al., 2005; Li and Widom, 2004) or actively by forces generated by polymerases and chromatin remodelers (Sirinakis et al., 2011; Yin et al., 1995). In addition, highly dynamic chromatin anchored to various subcellular structures is likely to experience tension. Nucleosomal DNA under tension has been proposed to unwrap in two major stages; the outer turn unwraps at low force followed by unwrapping of the inner turn at higher force (Brower-Toland et al., 2002; Mack et al., 2012; Mihardja et al., 2006). However,

previous mechanical studies relied on end-to-end distance detection of the DNA tethers, interpretation of which can be indirect, and is unable to report on local conformational changes of different parts of the nucleosome.

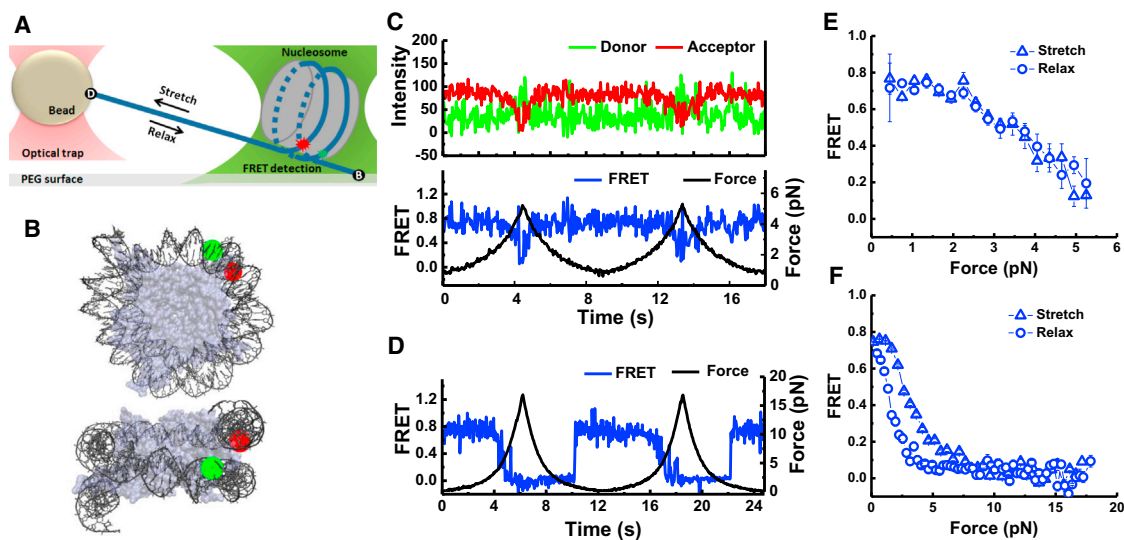
Understanding the physical basis of how DNA sequence and modifications affect nucleosome dynamics will help elucidate how genomic and epigenetic modifications regulate cellular functions. In the nucleosome, DNA of about one persistence length (147 bp) has to be bent and twisted to form ~1.7 turns around the histone octamer (Chua et al., 2012; Kulaeva et al., 2013; Luger et al., 1997). DNA sequence may affect the strength of DNA-histone interactions through formation of specific DNA-histone interactions or by affecting the static curvature, dynamic flexibility, permanent or dynamic twist (Widom, 2001). These mechanical properties of DNA are affected by sequence composition and a variety of modifications (Hagerman, 1988; Mirsaidov et al., 2009; Rief et al., 1999; Severin et al., 2011; Vafabakhsh and Ha, 2012; Widom, 2001). The DNA sequence has a profound effect on nucleosome positioning, structure, and stability (Chua et al., 2012; North et al., 2012; Tóth et al., 2013; Widom, 2001), but how it affects nucleosome dynamics is poorly understood.

Here, we employ a single-molecule assay which combines fluorescence with optical tweezers (Hohng et al., 2007; Maffeo et al., 2014; Zhou et al., 2011) to simultaneously manipulate an individual nucleosome under force and probe its local conformational transitions.

## RESULTS

### Probing Local Conformational Dynamics of the Nucleosome under Tension

In order to obtain clearly interpretable data on local nucleosome dynamics we chose the nucleosome positioning sequence 601 (Lowary and Widom, 1998), which has been used for previous high resolution single molecule studies (Bintu et al., 2011, 2012; Böhm et al., 2011; Brower-Toland et al., 2002; Deindl et al., 2013; Gansen et al., 2009; Hall et al., 2009; Hodges et al., 2009; Kruithof and van Noort, 2009; Mack et al., 2012; Mihardja et al., 2006; North et al., 2012; Sheinin et al., 2013; Shundrovsky et al., 2006; Sudhanshu et al., 2011; Tóth et al., 2013). A nucleosome was anchored to a PEG-coated glass surface on one end of the DNA and pulled via a  $\lambda$ -DNA tethered to the other end by an optical trap (Figure 1A). A fluorescence resonance



**Figure 1. Observation of Local Conformational Changes of Nucleosome under Tension**

(A) Experimental scheme: a nucleosome was immobilized on a microscope slide via a 14 bp dsDNA handle beyond the nucleosome core sequence. The other end was connected to a micron-diameter bead through a  $\lambda$ -DNA linker which was held in place by an optical trap which applies force. Local conformational changes were recorded by FRET between the donor (green) and the acceptor (red) on the DNA.

(B) Positions of donor and acceptor fluorophores in the ED1-labeling scheme superposed on two different views of the nucleosome structure (Protein Data Bank [PDB] file 3MVD).

(C and D) Single-molecule time traces of the ED1 construct recorded during stretching and relaxing at a stage speed of 455 nm/s at a set maximum force of  $\sim 6$  pN (C) and  $\sim 20$  pN (D): force (black), donor signal (green), acceptor signal (red), and FRET efficiency (blue).

(E and F) The average FRET versus force when the maximum force was set to  $\sim 6$  pN (E): average of 26 traces and  $\sim 20$  pN (F): average of 25 traces.

See also Figure S1.

energy transfer (FRET) dye pair, a donor and an acceptor, attached to various positions on the DNA enable the measurement of conformational changes of defined locality.

To probe unwrapping of the outer DNA turn, we constructed the ED1 (Entry-Dyad 1) labeling scheme consisting of a donor close to the dyad and an acceptor close to an entry. ED1 nucleosomes displayed a single high FRET population due to close proximity of the probes (Figure S1A) as expected from the nucleosome crystal structure (Makde et al., 2010) (Figure 1B). In the absence of force, FRET time traces were stable within our temporal resolution of 30 ms (Figure S1B). The same DNA reconstituted with the (H3/H4)<sub>2</sub> tetramer produced a very different distribution with low FRET values attributed to the tetrasome (Figure S1A).

We increased the applied force starting from a low value (typically between 0.4–1.0 pN) to a predetermined higher value and then returned it to the low value. FRET gradually decreased as the force increased followed by fast fluctuations and finally a sharp decrease in FRET (Figures 1C and 1D). Upon relaxation, the nucleosome reformed, retracing the dynamics observed during stretching if the force was held below 6 pN to limit the extent of unwrapping (Figures 1C and 1E) or displaying hysteresis when we extended the force range to 20 pN (Figures 1D and 1F). The initial gradual FRET decrease indicates that DNA unwraps steadily without going through a major energy barrier at low tension. The FRET fluctuation that follows likely represents a bistable hopping behavior reported previously (Mihardja et al., 2006). Subsequent stretching/relaxation cycles reproduced the same

behavior, suggesting that each cycle brings the nucleosome back to the initial state.

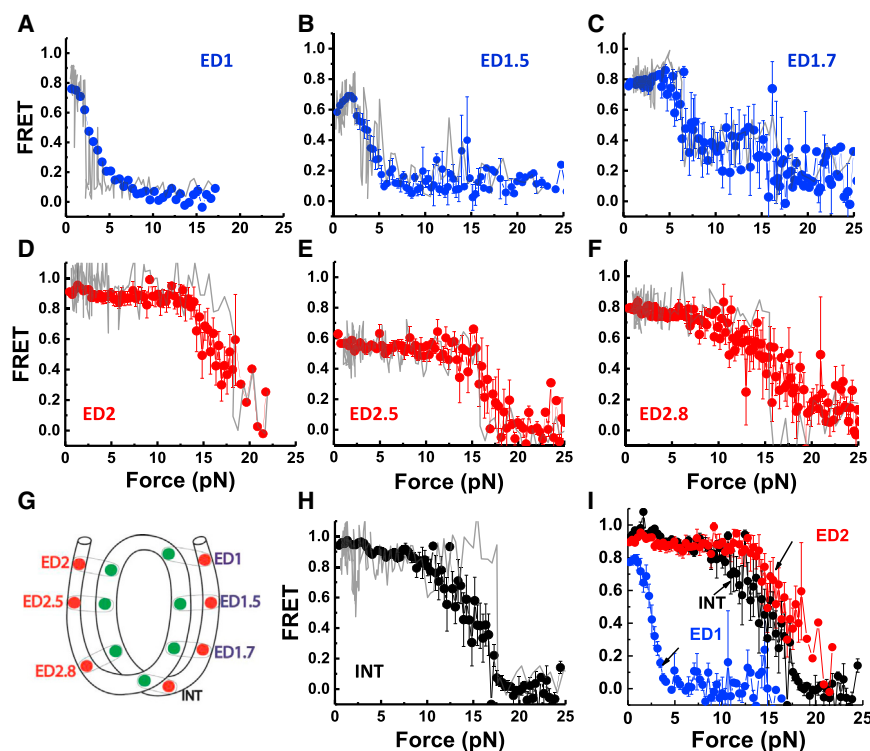
To probe inner turn unwrapping, we attached FRET probes to a region  $\sim 40$  bp from the dyad (INT) (Figure 2G). As with ED1, the INT nucleosome showed a single narrow FRET peak at zero force and was distinguishable from the tetrasome species that displayed a broad range of FRET (Figures S1C and S1D). At low forces, the INT nucleosome maintained a stable high FRET value with occasional hopping to an intermediate FRET state (Figure 2H). As the force increased to higher values (10–15 pN), FRET suddenly dropped to a final low value (Figure 2H). As an additional control, the INT-tetrasome showed a distinct FRET versus force stretching pattern, unraveling at much lower force (3–5 pN), thus confirming that INT nucleosome contained the histone octamer (Figure S1E).

Taken together, our nucleosome stretching data with probes at ED1 and INT positions are consistent with previous studies on the effect of force on global nucleosome dynamics (Brower-Toland et al., 2002; Kruithof and van Noort, 2009; Mihardja et al., 2006; Sheinin et al., 2013); the outer turn unwraps at low force (3–5 pN) and the inner turn unwraps at higher force (12–15 pN). In addition, ED1 probe allows us to observe gradual unwrapping before an abrupt transition of the initial DNA end segment at the low force range ( $< 3$  pN).

### Nucleosome Unwrapping Is Asymmetric

Previous investigations of nucleosome unwrapping (Brower-Toland et al., 2002; Kruithof and van Noort, 2009; Mack et al., 2012;





**Figure 2. Nucleosome Unwraps Directionally under Tension**

(A–H) FRET versus force during stretching for various FRET pairs spanning two sides of the nucleosome illustrated in (G) (see Figure S2 for labeling positions). Representative data for single cycles are shown in gray. The averaged curves are in blue for the weak side, in red for the strong side, and in black for the inner turn probes. Error bars are SEM of 25 traces for ED1 (A), 15 traces for ED1.5 (B), 8 traces for ED1.7 (C), 20 traces for ED2 (D), 7 traces for ED2.5 (E), 40 traces for ED2.8 (F), and 22 traces for INT (H).

(I) Overlay of ED1, ED2, and INT stretching curves. Substeps, which may arise from progressive unwrapping, could be seen for ED1.7 both in the averaged trace and in individual traces (three out of eight cycles).

See also Figure S2.

Mihardja et al., 2006; Sheinin et al., 2013) assumed that two nucleosomal DNA ends respond similarly to the applied force since unwrapping of the two DNA ends was not separately observable. Our assay, which is sensitive to local conformational changes, enables the examination of two sides separately by comparing the FRET-Force response on the two ends. We designed a construct termed ED2 with a FRET pair placed at the opposite entry/dyad region—the “left” end (Figures 2G and S2A). Surprisingly, the FRET-Force pattern of ED2 displayed a pattern very different from ED1 (Figure 2D). FRET remained stable at low forces and did not decrease until higher force (15–20 pN) was reached, in contrast to the decrease below 5 pN observed for ED1 on the “right” end. This result indicates that a significant asymmetry exists in the DNA unwrapping behavior.

We performed various control experiments to confirm the unwrapping asymmetry result and to rule out alternative explanations. First, additional constructs with probes at symmetric locations on the DNA handles outside the core sequence (Figure S3). Second, we swapped the orientation of surface tethering and pulling via the lambda DNA tether and found that the strong side unwrapped at high forces for both configurations (with essentially identical FRET versus force curves), ruling out surface tethering via a particular end as the reason for the asymmetry (Figure 3A). Third, replacing the first 10 bp of the left handle with the corresponding region on the right handle showed that the sequence difference just outside the core region is not responsible for the asymmetry (Figure 3B).

To examine if the observed asymmetry may be induced by position-specific perturbations caused by the fluorophores,

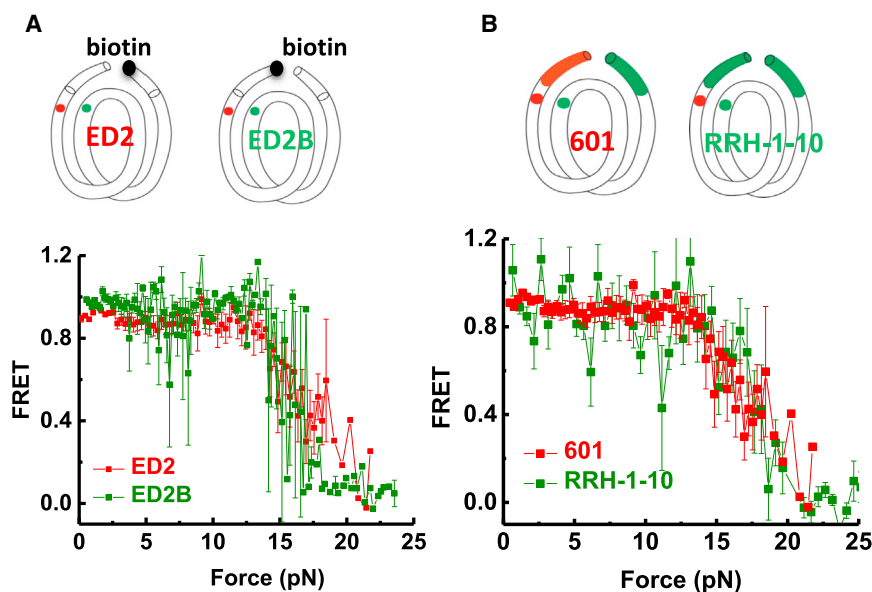
we designed four additional constructs for comparison of the two sides: ED1 versus ED2, ED1.5 versus ED2.5, and ED1.7 versus ED2.8 (Figures 2G and S2A). Generally, the force required for a significant FRET decrease was lower for ED1 (Figure 2A), ED1.5 (Figure 2B), and ED1.7 (Figure 2C) than for those labeled

at symmetrically related sites, ED2 (Figure 2D), ED2.5 (Figure 2E), and ED2.8 (Figure 2F), respectively, showing that the asymmetry is highly unlikely due to position-dependent perturbations by the fluorophores and indicating that one side of the nucleosome is indeed weaker than the other when the DNA is under tension. Strikingly, the force needed for a major unwrapping signal was larger for the ED2 end ( $16.8 \pm 0.4$  pN) than the DNA inner turn ( $14.7 \pm 0.5$  pN) (Figures 2D and 2H) (the errors represent the SEM). This effect was even clearer when the pulling rate was halved to 233 nm/s ( $14.2 \pm 0.5$  pN versus  $11.2 \pm 0.9$  pN; Figure S2). Thus, the data suggest that DNA unwrapping occurs directionally, starting from the “weak” end (ED1) at the lowest unwrapping force, followed by the inner turn, and then to the “strong” end (ED2). However, the small difference between the INT and ED2 unwrapping force would allow the inner turn to unwrap later than the strong end in some cases.

Such mechanical asymmetry may influence gene expression by affecting DNA exposure or transcriptional pausing. In fact, an in vitro transcription study (Bondarenko et al., 2006) observed that nucleosomes can form a polar barrier to transcriptional elongation. Specifically, our “strong” side (ED2) corresponds to the 601R transcription orientation where polymerases face a higher outer turn barrier (the +15 barrier).

### Unwrapping of the Nucleosome on One End Stabilizes the Other End

In the low force range, FRET of the strong outer turn ED2 is stable and remains unchanged until the final drop at high force ( $16.8 \pm 1.5$  pN) (Figure 2D). When the pulling rate is lowered 2-fold



**Figure 3. Unwrapping Force Is Not Affected by Pulling Configuration or Extra-Nucleosomal Handle Sequence**

(A) Switched pulling configurations for the same labeling position ED2. In the ED2 scheme, the 5' end of the bottom J strand (the right end) is biotinylated. In the ED2B scheme, the 5' end of the top I strand (the left end) is biotinylated. Averaged stretching traces for both ED2 pulling configurations show identical high force required for unwrapping (ED2: average of 20 traces, ED2B: average of 4 traces).

(B) Changing the handle sequence on the left side does not alter the high force range required to open nucleosomal DNA on this side. Averaged stretching curves show identical high force required for unwrapping for 601-ED2 (average of 20 traces) and RRH-1-10-ED2 (average of 15 traces).

See also Figure S3.

(Figure S2), we observed a small decrease in FRET followed by a FRET recovery in the low force range for some stretching traces. Therefore, we probed the earliest unwrapping process of the strong (ED2) side by moving the probes to either one (ED2-1) (Figure 4C) or twelve (ED2-12) (Figure S4A) nucleotides beyond the nucleosome core sequence on the strong side. At low forces, ED2-1 and ED2-12 probes on the strong side showed the same stretching pattern as ED1 probe on the weak side: FRET decreased gradually at low force followed by fluctuations at 3–6 pN (Figures 4A and S4). However, on the weak side the FRET dropped entirely after 6 pN, while on the strong side, FRET recovered and did not fully drop until much higher force was reached. Our force-fluorescence spectroscopy approach allows detection of unwrapping/rewrapping of a specific side. In contrast, in previous studies measuring the overall end-to-end distance (Mihardja et al., 2006; Sheinin et al., 2013), simultaneous rewrapping of the strong end and unwrapping of the weak end may not have given a detectable change in overall length. Coordination in FRET-force patterns of ED1 and ED2-1 indicate that two extreme ends of the nucleosome are slightly unwrapped at low forces but once the weak end significantly unwraps, the strong end rewraps and stays stable until much higher forces are applied.

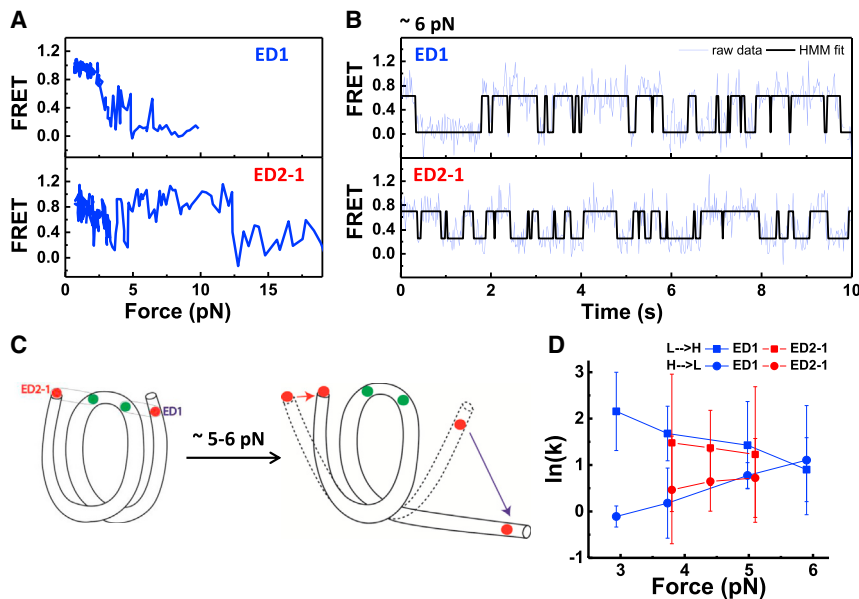
At constant forces, FRET time traces of both ED1 and ED2-1 constructs (Figure 4B) showed two-state hopping between wrapped and partially unwrapped state, respectively. Hidden Markov Modeling (McKinney et al., 2006) was used to determine the transition rates between the two states (Figure 4D). As the force increased, the unwrapping rate increased and the wrapping rate decreased, consistent with a previous report (Mihardja et al., 2006), and the rates on the two DNA ends were similar. Our observation that the two ends of the nucleosome are orchestrated such that the opening of one end helps stabilize the other end raises the possibility that even relatively small asymmetry between the two sides may result in one side winning reliably (cartoon in Figure 4C).

### Asymmetry of Nucleosome Unwrapping Is Directed by DNA Local Flexibility

We propose that the observed asymmetry in mechanical stability originates from the DNA sequence differences between the two sides of the 601 sequence for the following reasons. First, the protein core structure is symmetric around the dyad axis (Chua et al., 2012; Luger et al., 1997) whereas the DNA sequence is nonpalindromic. Second, unzipping of the nucleosomal DNA under certain experimental configurations shows a higher off-dyad barrier on one side (“strong” side in our study) than the other (Hall et al., 2009). Third, symmetrization of certain sequence features can affect the overall thermodynamic stability as measured by salt titration (Chua et al., 2012).

Since the unwrapping asymmetry is observed for the outer turn, we first symmetrized DNA content at the entry regions by replacing the AT-rich region (nucleotide 8–24 from the right end) on the weak side with the corresponding GC-rich segment on the strong side (Figure S5A). This construct, termed LL8–24, exhibited the same asymmetry as the 601 nucleosome (Figure S5C), ruling out the differences in AT/GC-content of the entry region as the source of asymmetry.

Because DNA has to be bent and deformed to wrap around the histone octamer, the intrinsic DNA flexibility may influence DNA-histone binding affinity (Lowary and Widom, 1998; Widom, 2001). Therefore, we hypothesized that the more flexible sequences would unravel at higher forces by better tolerating the sharply bent DNA conformation. To test this hypothesis, we examined the relative flexibility of the two 73 bp DNA fragments flanking the dyad in the 601 sequence using a single molecule DNA cyclization assay (Vafabakhsh and Ha, 2012). The “strong” side (LH for left half) yielded a cyclization time of 26 min while the “weak” side (RH for right half) took 189 min to cyclize, indicating that the left side of the 601 is more flexible than the right side by a factor of 7 according to our measurement (Figure 5C). Thus, the asymmetry in DNA flexibility appears to correlate with asymmetric unwrapping—the more flexible DNA side unwraps at



**Figure 4. Coordinated Dynamics of the Two Nucleosomal DNA Ends**

(A) Representative single-molecule stretching traces of ED1 and ED2-1 as indicated in (C).

(B) Representative time traces of FRET efficiency at a constant force of 6 pN, showing hopping between high and low FRET states. Fits from Hidden Markov modeling are overlaid.

(C) Illustration of how major unwrapping of one side of the nucleosome facilitates rewinding on the other end. Initially, two extreme ends of the nucleosome synchronously unwrap and rewrap at forces below ~5 pN (dashed shape). Once the ED1 side majorly unwraps (blue arrow), this facilitates the rewinding of the ED2 side (red arrow).

(D) Rates of transition between high and low FRET states versus force. Unwrapping rates (high to low FRET transitions) in circles and rewrapping rates (low to high FRET transitions) in squares.

See also Figure S4.

higher force and vice versa. Here, “flexibility” is an operational definition equivalent to “cyclizability” in our assay because we do not yet know whether a static bend or dynamic flexibility (represented by lower bending energy) determines the apparent flexibility.

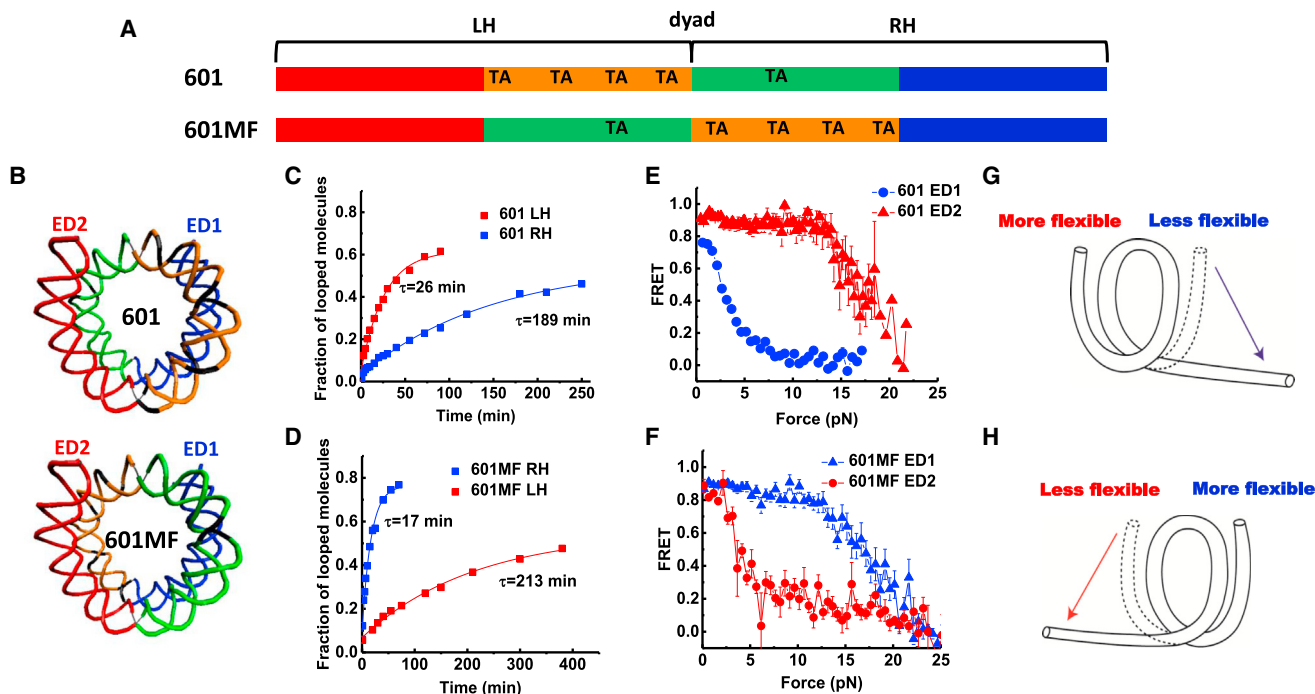
In order to test the correlation further, we modified the 601 sequence so as to locally switch the DNA flexibility on the two sides by flipping the middle 73 bp (601MF) (Figures 5A and 5B). The single molecule cyclization showed that the right side of 601MF has now become more flexible (17 min looping time) than the left side (213 min looping time) by a factor of 12 (Figure 5D), reversing the relation found in the original 601 sequence, and correspondingly, the left side of the 601MF nucleosome (now containing stiffer DNA sequence) unwrapped at a lower force than the right side (Figure 5F). This implies that the direction of outer turn unwrapping can be controlled by the relative flexibility of internal regions of DNA such that the nucleosome first unwraps from the DNA side connected to a less flexible inner turn DNA (Figures 5G and 5H).

We further tested how nucleosome unwrapping is affected when the DNA sequence is similar in flexibility on both sides. We were guided by the 10 bp TA steps rule suggested by Widom (Lowary and Widom, 1998; Widom, 2001) to construct this DNA sequence. Chua et al. (2012) confirmed by crystallography that TA dinucleotides accommodate the highest degree of distortion of the DNA structure within nucleosome. The 601 sequence is nonpalindromic with 10 bp TA steps situated only on the left (strong) side. Therefore, we pseudo-symmetrized the flexibility of the sequence by adding three copies of TA dinucleotides spaced 10 bp apart to the right (weak) side (601RTA) (Figures 6A and 6B). The resulting 601RTA right half (RH) became more flexible (cyclization time decreased from 189 to 63 min) (Figure 6C) and closer to the left half (26 min). We ensured that the nucleosome positioning is maintained on all three sequences (601, 601MF, and 601RTA), as nucleosomes reconstituted

from all three sequences show the same electrophoretic mobility on a 5% native PAGE gel and displayed similar single-molecule FRET histograms (Figure S6). Strikingly, instead of one side winning the match every time, which side unwraps at low forces became stochastic (Figure 6F). The fraction of traces unwrapped at low force and high force was 37% and 67% for the left half (601RTA-ED2) and 44% and 56% for the right half (601RTA-ED1), respectively (Figures 6G and 6H). Averaging over all stretching traces produced almost identical FRET-force patterns for these two constructs (Figure 6D). These results imply that when the flexibility of DNA on the two sides of the nucleosome is similar, each side of the nucleosome unwraps stochastically at either low force or high force (Figure 6E).

### Monte Carlo Simulation of Asymmetric Unwrapping of Nucleosomal DNA

In order to model the asymmetric nucleosome dynamics under tension, we adopted a continuum model of symmetric nucleosomal DNA unwrapping developed by Sudhanshu et al. (2011) and extended it to a more general, asymmetric case where  $m$  and  $n$  base pairs can be unwrapped from the weak and strong side, respectively. The only modification to the energy function used by Sudhanshu et al. (2011) was a reduction in the binding energy of the inner quarter of the weak side (see Supplemental Information for details). With this energy function, we performed Monte Carlo simulations starting from 0.1 pN and increasing the force in 0.1 pN increments every 2000 time steps until 10 pN of force was reached. Four representative trajectories of  $m$  and  $n$  values, the number of base pairs unwrapped from the weak and strong sides, respectively, are shown in Figure 7 (blue for  $m$  and red for  $n$ ). At ~3–5 pN of force, we observed major unwrapping of the weak side ( $m$  values reaching around 65 bp). In three cases (Figures 7A, 7B, and 7D), initial unwrapping of the strong side (transient increase in  $n$ , i.e., unwrapping of the strong side) precedes rewrapping of the strong side and major unwrapping of the weak side.



**Figure 5. Asymmetric Nucleosome Unwrapping Controlled by DNA Local Flexibility**

(A) Variations of the 601 sequence where the inner quarters are colored in orange and green and the outer quarters are colored in red and blue. TA steps are indicated.

(B) Nucleosomal DNA structures are shown in the same color scheme with corresponding scheme of the sequence.

(C and D) Single exponential fits to the looped DNA fraction versus time yield the average looping time  $\tau$  measured using single DNA cyclization assay for the 73 bp left or right halves (LH and RH, respectively).

(E and F) Averaged stretching time traces of FRET efficiency versus force for nucleosomes in ED1 and ED2 labeling schemes. Error bars denote SEM of 25 traces for 601 ED1, 15 traces for 601 ED2, 29 traces for 601MF ED1, 19 traces for 601MF ED2.

(G and H) Illustrations of the relationship between the direction of nucleosome unwrapping and the DNA flexibility of the two halves of the nucleosomal DNA sequence. The nucleosome unwraps from the stiffer side (single-headed arrows) if the DNA flexibility differs significantly between the two sides.

See also Figure S5.

Figure 7E shows an example trajectory in  $(m, n)$  space (corresponding to Figure 7B). A transient unwrapping of the strong side is seen in the force range 3–4 pN before the systems moves to the asymmetrically unwrapped state.

This simple model and simulation capture two important aspects of our data. First, asymmetric unwrapping can be obtained even when only the inner quarters are different in binding energy (presumably arising from differences in DNA flexibility where less flexible sequence has less binding energy). Second, a transient unwrapping of the strong side is often observed, and this is followed by rewrapping of the strong side and major unwrapping of the weak side in a coordinated fashion. Furthermore, our data and simulation suggest that the force-induced extension changes observed in previous studies at low forces and interpreted as symmetric unwrapping of the outer turns from both ends may need to be reinterpreted as asymmetric unraveling of the weak side only.

## DISCUSSION

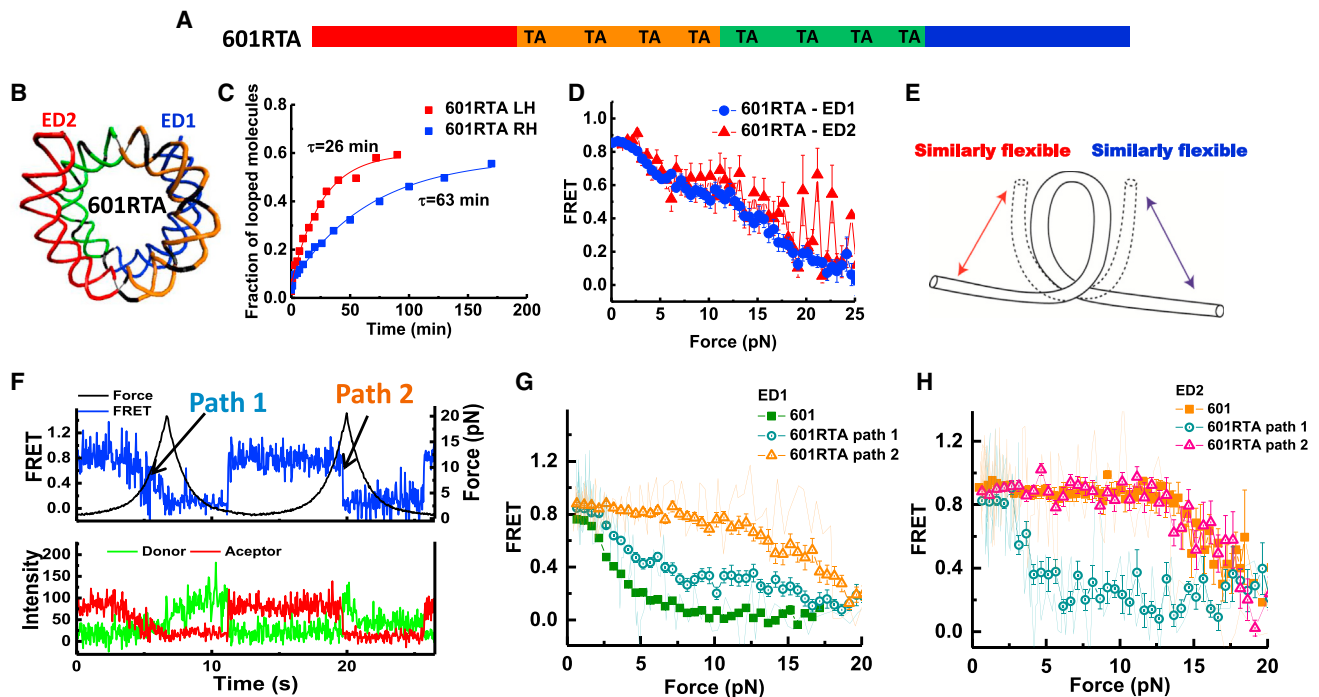
Genetic information buried in nucleosome is made accessible for replication, transcription, repair, and remodeling by partial

unwrapping of nucleosomes (Bowman, 2010; Gansen et al., 2009; Hodges et al., 2009; Kulaeva et al., 2013; Li et al., 2005; Li and Widom, 2004; Maher et al., 2013; North et al., 2012; Tims et al., 2011). Our results provide the first demonstration of how the local flexibility of DNA governs the mechanical stability of the nucleosome and accessibility of nucleosomal DNA and may be generalizable as a principal mechanism for regulation of DNA metabolism by nucleosomal DNA sequence and modifications.

The correlation that the more flexible the DNA sequence is, the more stable it stays bound to the histone core may aid the prediction of nucleosome positions imposed by DNA sequence. We found that this relation holds not only for DNA sequences but also for DNA modifications such as DNA mismatches, 5-methylcytosine and 5-formylcytosine (T.T.M.N., Q.Z., J. Yoo, Q. Dai, A. Aksimentiev, C. He, and T.H., unpublished data).

Stabilization of one nucleosomal DNA end upon the major opening of the other end may play a role in nucleosome integrity maintenance during transcription and nucleosome remodeling because both in vivo and in vitro studies suggest that a high fraction of nucleosomes survive after being transcribed (Bintu et al., 2011; Workman, 2006) and remodeled (Shundrovsky





**Figure 6. Stochastic Unwrapping of Nucleosome on the Sequence with Similar Flexibility on Two Sides**

(A) Scheme of the 601RTA sequence which is derived from the 601 sequence by substitution of three dinucleotides on the right side by three TA steps. (B) Nucleosomal DNA structures are shown in the same color scheme with the scheme of the sequence. (C) Single exponential fits to the looped DNA fraction versus time yield the average looping time  $\tau$  measured using single DNA cyclization assay for the 73 bp left or right halves (LH and RH, respectively) for the 601RTA sequence. (D) Averaged stretching time traces of FRET efficiency versus force for nucleosomes in ED1 (average of 57 traces) and ED2 (average of 7 traces) labeling schemes for the 601 RTA sequence. Error bars denote SEM. (E) A cartoon illustrating stochastic unwrapping of nucleosome from either side when the DNA flexibility on the two sides is made similar on the 601RTA sequence. (F) Representative single-molecule fluorescence-force time trace for 601-RTA nucleosome reconstituted with the ED1 labeling scheme. Two unwrapping paths are shown. Path 1 is gradual FRET decrease at low force (similar to original weak side), while path 2 is sudden FRET decrease at high force (similar to original strong side). (G and H) Averaged FRET versus force stretching curves for 601-RTA-ED1 (25 traces for path 1 and 32 traces for path 2) nucleosomes (G) and 601-RTA-ED2 (four traces for path 1 and three traces for path 2) nucleosomes (H), comparing to that of ED1 and ED2 of the 601 sequence. Representative single-molecule stretching traces are shown in lighter color lines. See also Figure S6.

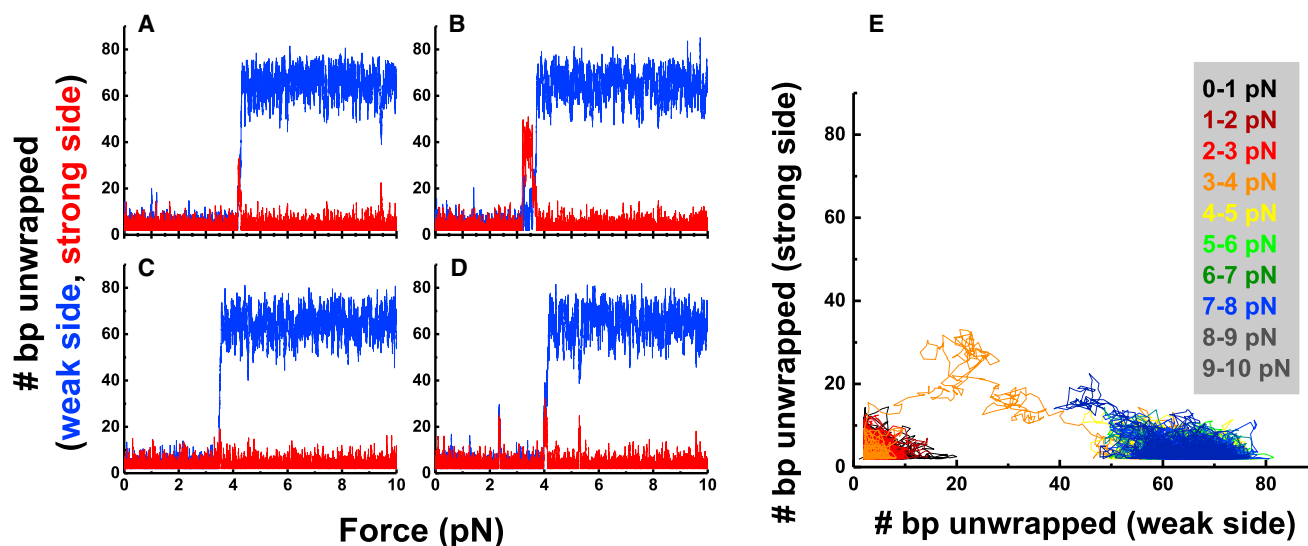
et al., 2006). It is also possible that such orchestration between the two nucleosome ends may help stabilize one H2A/H2B dimer during the exchange or modification of the other dimer. For example, SWR-C/SWR-1 deposits H2A.Z into only one site at a time, not both (Yen et al., 2013).

Our Monte Carlo simulations could reproduce key features of asymmetric unwrapping and coordinated dynamics of two DNA ends (Figure 7). Nevertheless, this model ignores many structural details and represents DNA sequences with the resolution of 36 bp, a quarter of the nucleosomal DNA. Other properties of the nucleosome yet to be explored may make additional contributions to the coordination of DNA ends: (1) the proposed electrostatic repulsion between two DNA turns (Mollazadeh-Beidokhti et al., 2012) where upon force-induced undocking of one end, the resulting loss of the electrostatic repulsion stabilizes the other end, (2) DNA allostery (Kim et al., 2013), and (3) the deformation of the histone octamer during unwrapping which may change charge distribution and/or contribute to the allosteric

coupling. Histone deformation was suggested to govern salt-induced nucleosome dissociation (Böhm et al., 2011) and may also be involved in nucleosome remodeling by IWSI remodelers (Deindl et al., 2013). In our experiments at low tension, in addition to the early unwrapping of extreme DNA ends probed by ED1 and ED2-1, the FRET probes at ED1.5 (Figure 2B) and ED1.7 (Figure 2C) displayed an increase in FRET as a first response to applied force before a decrease, indicating possible partial DNA tightening mediated by twisting of the H<sub>2</sub>A/H<sub>2</sub>B dimer on the weak side.

We observed that nucleosomal DNA unwraps directionally under tension not only for the 601 sequence but also for the derivatives of the 601 sequence. Asymmetric unwrapping is likely to be generalizable to other sequences since the coordination of two ends would allow the system to amplify even a small difference in flexibility to cause a large asymmetry in mechanical stability.

Directionality of transcription can be ensured by the suppression of cryptic antisense (Gorman et al., 2010) through



**Figure 7. Monte Carlo Simulation of Nucleosome Unwrapping**

(A–D) Representative Monte Carlo simulation records show the number of base pairs unwrapped from the weak side (blue) and the strong side (red) as the force increased from 0.1 pN to 10 pN.

(E) A 2D representation of unwrapping trajectory shown in (B). Different portions of the trajectory at difference forces are shown in different colors as indicated.

epigenetic regulation and RNA degradation (Richard and Manley, 2013). Our results linking sequence-dependent flexibility to mechanical stability of the nucleosome suggest another mechanism to maintain transcriptional direction—the possibility that nature selects for lower flexibility DNA sequences within the first half of nucleosomes in the direction of transcription. In this scenario, RNA polymerase would have greater initial access to the DNA template if it enters the nucleosomal DNA from the “weak” side and would only pause when it reaches the nucleosomal dyad (Churchman and Weissman, 2011; Hodges et al., 2009). We are currently investigating DNA flexibility on a genomic scale combining sequencing and single molecule cyclization to test this possibility.

## EXPERIMENTAL PROCEDURES

For additional details, see the [Extended Experimental Procedures](#).

### Preparation of DNA Constructs

We used PCR to amplify 181 bp ds DNA from templates that contain 147 bp 601 positioning sequence, flanked by a 14 bp linker to biotin and 20 bp spacer connected to the 12 nt COS overhang. The construct was tethered to the surface via biotin and the COS overhang was used to anneal the template to  $\lambda$  DNA. PCR primer oligonucleotides were designed for various templates and synthesized by Integrated DNA Technologies. The forward primer contains an amino modification (5AmMC6T) at a designated location and a biotin at the 5' end. The reverse primer contains the same amino modification and an abasic site to create the COS overhang. The forward and reverse primers were labeled with Cy3 and Cy5 dyes, respectively, according to Roy et al. (2008) and HPLC-purified when necessary to bring the labeling efficiency to >90%.

### Nucleosome Reconstitution

PCR-amplified 601 templates were reconstituted with *Xenopus laevis* recombinant histone octamer (purchased from Colorado State University) by salt-

dialysis (Dyer et al., 2004). Reconstituted nucleosomes were stored at 4°C in the dark typically at concentrations of 100–200 nM and used within 2 weeks. The efficiency of nucleosome reconstitution was measured by 5% native PAGE gel electrophoresis.

### Annealing Nucleosome to $\lambda$ DNA

The nucleosome was annealed to  $\lambda$  DNA and an oligonucleotide containing digoxigenin. First,  $\lambda$  DNA (NEB) at 16 nM was heated in the presence of 120 mM NaCl and 1.2 mM  $MgCl_2$  at 80°C for 10 min and then placed on ice for 5 min. Nucleosomes and BSA were added to the  $\lambda$  DNA at a final concentration of 8 nM and 0.1 mg/ml, respectively. The mixture was incubated with rotation in the dark at room temperature for 15 min and then for an additional 2–3 hr at 4°C. DIG oligo (see [DNA Sequences](#) in the [Supplemental Information](#)) was added to a final concentration of 200 nM and then incubated with rotation at 4°C for 1–2 hr. Samples were stored at 4°C in the dark and could be used for data acquisition for up to 2 weeks.

### Sample Assembly

To eliminate nonspecific surface binding, a coverslip surface was coated with polyethyleneglycol (PEG) (mixture of mPEG-SVA and Biotin-PEG-SVA, Laysan Bio) according to Roy et al. (2008). After forming an imaging chamber using the PEG coated coverslip and glass microscope slide, it was further incubated in blocking buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mg/ml BSA [NEB], 1 mg/ml tRNA [Ambion]) for 1 hr. The nucleosome sample was diluted to 10 pM in a nucleosome dilution buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM  $MgCl_2$ ) and immobilized on the surface via biotin-neutravidin interaction. Next, 1  $\mu$ m anti-digoxigenin-coated polystyrene beads (Polysciences) diluted in nucleosome dilution buffer were added to the imaging chamber for ~30 min for attachment of beads to the free end of each tether. Finally, imaging buffer (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM  $MgCl_2$ , 0.5 mg/ml BSA [NEB], 0.5 mg/ml tRNA [Ambion], 0.1% v/v Tween-20 [Sigma], 0.5% w/v D-Glucose [Sigma], 165 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [Roche], and 3 mM Trolox [Sigma]) was added for data acquisition.

### Fluorescence-Force Spectroscopy

We recently developed an instrument combining optical trap with fluorescence detection to monitor conformational changes of biomolecular systems under applied force (Hohng et al., 2007). The full details of this instrument can be

found in our recent review (Zhou et al., 2010). Briefly, an optical trap was formed by an infrared laser (1,064 nm, 800 mW, EXLSR-1064-800-CDRH, Spectra-Physics) through the back port of the microscope (Olympus) by expanding the laser beam 8-fold using two telescopes and focusing on the sample plane with a 100 $\times$  oil immersion objective (Olympus). Force was applied on the sample tethers by moving the microscope slide using a piezo stage (Physik Instrument). Applied force was determined by position detection of the tethered beads using a QPD (UDT/SPOT/9DMI) and stiffness calibration as described (Hohng et al., 2007). The confocal excitation laser (532 nm, 30 mW, World StarTech) was coupled through the right port of the microscope. The excitation laser was scanned by a piezo-controlled steering mirror (S-334K.2SL, Physik Instrument). The fluorescence emission was filtered from the infrared laser by a band pass filter (HQ580/60 m, Chroma) and separated from excitation by a dichroic mirror (HQ680/60 m, Chroma) before detection by two avalanche photodiodes.

### Data Acquisition

Single molecule data acquisition was performed according to Hohng et al. (2007). In summary, after a bead was trapped, the origin of the tether was determined by stretching the tether in two opposite directions along both x and y axis. Then the confocal laser was scanned to locate the fluorescence spot on the tether after separating the trapped bead from its origin by 14  $\mu$ m. Unless specified otherwise, the nucleosome unwrapping experiment was carried out by moving the stage between 14  $\mu$ m and 16.8–17.2  $\mu$ m at the speed of 455 nm/s<sup>-1</sup>. The confocal excitation was scanned concurrently with the stage movement. Fluorescence emission was detected for 20 ms after each step in stage movement. Force-fluorescence data was obtained in the imaging buffer (50 mM Tris-HCl pH 8, 50 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.5 mg/ml BSA [NEB], 0.5 mg/ml tRNA [Ambion], 0.1% v/v Tween-20 [Sigma], 0.5% w/v D-Glucose [Sigma], 165 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [Roche] and 3 mM Trolox [Sigma]).

### Single-Molecule DNA Cyclization Assay

A single-molecule DNA cyclization assay was recently developed in our laboratory to quantify the flexibility of a short double stranded DNAs (<100 bps) (Vafabakhsh and Ha, 2012). A total of six 601 DNA fragment regions listed in DNA Templates and Labeling Schemes in the Supplemental Information are generated by slow annealing (90°C for 10 min) of appropriate oligonucleotides (see DNA Sequences in the Supplemental Information) followed by slow cooling to room temperature over 4 hr. DNA fragments were immobilized on a PEG-coated microscope slide via biotin-neutravidin linkage. A FRET pair (Cy3 and Cy5) was incorporated at the two 10 nt long 5' overhangs that are complementary to each other so that loop formation via annealing of the two overhangs was detected as a FRET increase. Data acquisition was performed in a buffered solution (10 mM Tris-HCl pH 8.0, 1 M NaCl, 0.5% w/v D-Glucose [Sigma], 165 U/ml glucose oxidase [Sigma], 2170 U/ml catalase [Roche], and 3 mM Trolox [Sigma]). Time courses of generation of high FRET population allowed us to quantify the fraction of looped molecules versus time after the high salt buffer was introduced to the chamber containing low salt buffer (10 mM NaCl) of otherwise identical composition. Here, the rate of loop formation was used as a measure of DNA flexibility. The faster the looping occurs, the more flexible the sequence is.

All single molecule measurements were performed at  $\sim$ 22°C.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.001>.

### AUTHOR CONTRIBUTIONS

T.N. designed the research, prepared the samples, conducted all force-fluorescence experiments and part of DNA looping experiments, analyzed and interpreted the data, and wrote the manuscript. Q.Z. conducted part of looping experiments. R.Z. contributed to the early stage of instrumentation. J.Y. over-

saw the initial stage of the project and revised the manuscript. T. H. supervised the project, performed Monte Carlo simulations, and revised the manuscript. All authors discussed the results and were given opportunities to revise the manuscript.

### ACKNOWLEDGMENTS

This work was supported by the NIH (GM065367) and by the National Science Foundation Physics Frontiers Center program (PHY 0822613 and PHY 1430124). We thank A.J. Spakowitz for providing the energy function used for Monte Carlo simulations. T.H. is an investigator with the Howard Hughes Medical Institute.

Received: August 23, 2014

Revised: October 7, 2014

Accepted: January 17, 2015

Published: March 12, 2015

### REFERENCES

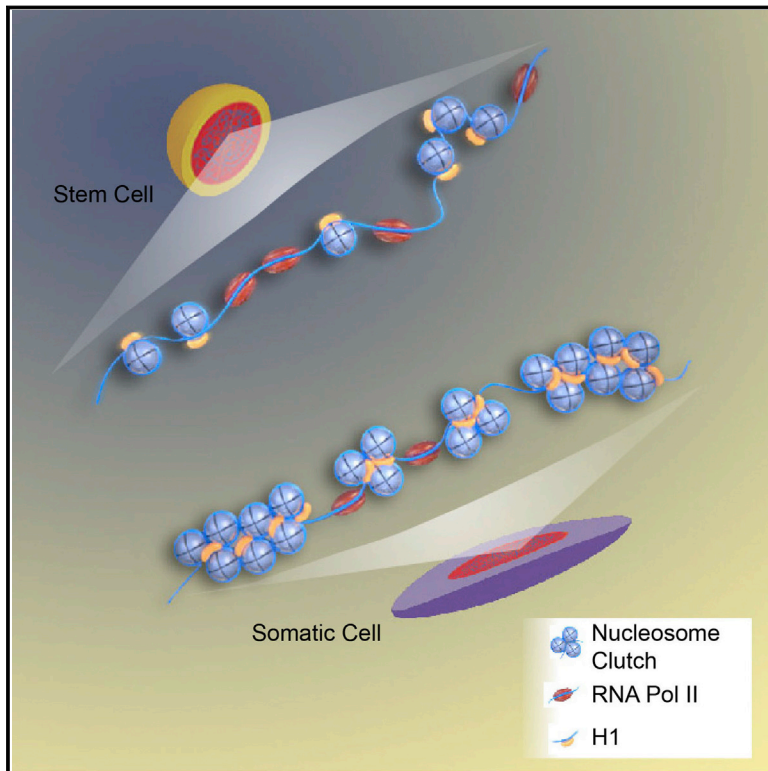
- Andrews, A.J., and Luger, K. (2011). Nucleosome structure(s) and stability: variations on a theme. *Ann. Rev. Biophys.* 40, 99–117.
- Bintu, L., Kopaczynska, M., Hodges, C., Lubkowska, L., Kashlev, M., and Bustamante, C. (2011). The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat. Struct. Mol. Biol.* 18, 1394–1399.
- Bintu, L., Ishibashi, T., Dangkulwanich, M., Wu, Y.-Y., Lubkowska, L., Kashlev, M., and Bustamante, C. (2012). Nucleosomal elements that control the topography of the barrier to transcription. *Cell* 151, 738–749.
- Böhm, V., Hieb, A.R., Andrews, A.J., Gansen, A., Rocker, A., Tóth, K., Luger, K., and Langowski, J. (2011). Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Res.* 39, 3093–3102.
- Bondarenko, V.A., Steele, L.M., Ujvári, A., Gaykalova, D.A., Kulaeva, O.I., Polikanov, Y.S., Luse, D.S., and Studitsky, V.M. (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol. Cell* 24, 469–479.
- Bowman, G.D. (2010). Mechanisms of ATP-dependent nucleosome sliding. *Curr. Opin. Struct. Biol.* 20, 73–81.
- Brower-Toland, B.D., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L., and Wang, M.D. (2002). Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA. *Proc. Natl. Acad. Sci. USA* 99, 1960–1965.
- Chua, E.Y.D., Vasudevan, D., Davey, G.E., Wu, B., and Davey, C.A. (2012). The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res.* 40, 6338–6352.
- Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368–373.
- Deindl, S., Hwang, W.L., Hota, S.K., Blosser, T.R., Prasad, P., Bartholomew, B., and Zhuang, X. (2013). ISWI remodelers slide nucleosomes with coordinated multi-base-pair entry steps and single-base-pair exit steps. *Cell* 152, 442–452.
- Dyer, P.N., Edayathumangalam, R.S., White, C.L., Bao, Y., Chakravarthy, S., Muthurajan, U.M., and Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* 375, 23–44.
- Gansen, A., Valeri, A., Hauger, F., Felekyan, S., Kalinin, S., Tóth, K., Langowski, J., and Seidel, C.A. (2009). Nucleosome disassembly intermediates characterized by single-molecule FRET. *Proc. Natl. Acad. Sci. USA* 106, 15308–15313.
- Gorman, J., Plys, A.J., Visnapuu, M.-L., Alani, E., and Greene, E.C. (2010). Visualizing one-dimensional diffusion of eukaryotic DNA repair factors along a chromatin lattice. *Nat. Struct. Mol. Biol.* 17, 932–938.
- Hagerman, P.J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. Chem.* 17, 265–286.

- Hall, M.A., Shundrovsky, A., Bai, L., Fulbright, R.M., Lis, J.T., and Wang, M.D. (2009). High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat. Struct. Mol. Biol.* **16**, 124–129.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**, 626–628.
- Hohng, S., Zhou, R., Nahas, M.K., Yu, J., Schulten, K., Lilley, D.M.J., and Ha, T. (2007). Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the holliday junction. *Science* **318**, 279–283.
- Kim, S., Brostr mer, E., Xing, D., Jin, J., Chong, S., Ge, H., Wang, S., Gu, C., Yang, L., Gao, Y.Q., et al. (2013). Probing allostery through DNA. *Science* **339**, 816–819.
- Koopmans, W.J.A., Brehm, A., Logie, C., Schmidt, T., and van Noort, J. (2007). Single-pair FRET microscopy reveals mononucleosome dynamics. *J. Fluoresc.* **17**, 785–795.
- Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868–871.
- Kruthof, M., and van Noort, J. (2009). Hidden Markov analysis of nucleosome unwrapping under force. *Biophys. J.* **96**, 3708–3715.
- Kulaeva, O.I., Malyuchenko, N.V., Nikitin, D.V., Demidenko, A.V., Chertkov, O.V., Efimova, N.S., Kirpichnikov, M.P., and Studitsky, V.M. (2013). Molecular mechanisms of transcription through a nucleosome by RNA polymerase II. *Mol. Biol.* **47**, 655–667.
- Li, G., and Widom, J. (2004). Nucleosomes facilitate their own invasion. *Nat. Struct. Mol. Biol.* **11**, 763–769.
- Li, G., Levitus, M., Bustamante, C., and Widom, J. (2005). Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12**, 46–53.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* **128**, 707–719.
- Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42.
- Luger, K., M der, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8   resolution. *Nature* **389**, 251–260.
- Mack, A.H., Schlingman, D.J., Ilagan, R.P., Regan, L., and Mochrie, S.G.J. (2012). Kinetics and thermodynamics of phenotype: unwinding and rewinding the nucleosome. *J. Mol. Biol.* **423**, 687–701.
- Maffeo, C., Ngo, T.T.M., Ha, T., and Aksimentiev, A. (2014). A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.* **10**, 2891–2896.
- Maher, R.L., Prasad, A., Rizvanova, O., Wallace, S.S., and Pederson, D.S. (2013). Contribution of DNA unwrapping from histone octamers to the repair of oxidatively damaged DNA in nucleosomes. *DNA Repair (Amst.)* **12**, 964–971.
- Makde, R.D., England, J.R., Yennawar, H.P., and Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**, 562–566.
- McKinney, S.A., Joo, C., and Ha, T. (2006). Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951.
- Mihardja, S., Spakowitz, A.J., Zhang, Y., and Bustamante, C. (2006). Effect of force on mononucleosomal dynamics. *Proc. Natl. Acad. Sci. USA* **103**, 15871–15876.
- Mirsaidov, U., Timp, W., Zou, X., Dimitrov, V., Schulten, K., Feinberg, A.P., and Timp, G. (2009). Nanoelectromechanics of methylated DNA in a synthetic nanopore. *Biophys. J.* **96**, L32–L34.
- Mollazadeh-Beidokhti, L., Mohammad-Rafiee, F., and Schiessel, H. (2012). Nucleosome dynamics between tension-induced states. *Biophys. J.* **102**, 2235–2240.
- Nag, R., and Smerdon, M.J. (2009). Altering the chromatin landscape for nucleotide excision repair. *Mutat. Res.* **682**, 13–20.
- North, J.A., Shimko, J.C., Javaid, S., Mooney, A.M., Shoffner, M.A., Rose, S.D., Bundschuh, R., Fishel, R., Ottesen, J.J., and Poirier, M.G. (2012). Regulation of the nucleosome unwrapping rate controls DNA accessibility. *Nucleic Acids Res.* **40**, 10215–10227.
- Richard, P., and Manley, J.L. (2013). How bidirectional becomes unidirectional. *Nat. Struct. Mol. Biol.* **20**, 1022–1024.
- Rief, M., Clausen-Schaumann, H., and Gaub, H.E. (1999). Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* **6**, 346–349.
- Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516.
- Severin, P.M.D., Zou, X., Gaub, H.E., and Schulten, K. (2011). Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.* **39**, 8740–8751.
- Sheinin, M.Y., Li, M., Soltani, M., Luger, K., and Wang, M.D. (2013). Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat. Commun.* **4**, 2579.
- Shundrovsky, A., Smith, C.L., Lis, J.T., Peterson, C.L., and Wang, M.D. (2006). Probing SWI/SNF remodeling of the nucleosome by unzipping single DNA molecules. *Nat. Struct. Mol. Biol.* **13**, 549–554.
- Sirinakis, G., Clapier, C.R., Gao, Y., Viswanathan, R., Cairns, B.R., and Zhang, Y. (2011). The RSC chromatin remodelling ATPase translocates DNA with high force and small step size. *EMBO J.* **30**, 2364–2372.
- Sudhanshu, B., Mihardja, S., Koslover, E.F., Mehraeen, S., Bustamante, C., and Spakowitz, A.J. (2011). Tension-dependent structural deformation alters single-molecule transition kinetics. *Proc. Natl. Acad. Sci. USA* **108**, 1885–1890.
- Tims, H.S., Gurunathan, K., Levitus, M., and Widom, J. (2011). Dynamics of nucleosome invasion by DNA binding proteins. *J. Mol. Biol.* **411**, 430–448.
- T th, K., B hm, V., Sellmann, C., Danner, M., Hanne, J., Berg, M., Barz, I., Gansen, A., and Langowski, J. (2013). Histone- and DNA sequence-dependent stability of nucleosomes studied by single-pair FRET. *Cytometry A* **83**, 839–846.
- Vafabakhsh, R., and Ha, T. (2012). Extreme bendability of DNA less than 100 base pairs long revealed by single-molecule cyclization. *Science* **337**, 1097–1101.
- Widom, J. (2001). Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324.
- Workman, J.L. (2006). Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009–2017.
- Yen, K., Vinayachandran, V., and Pugh, B.F. (2013). SWR-C and INO80 chromatin remodelers recognize nucleosome-free regions near +1 nucleosomes. *Cell* **154**, 1246–1256.
- Yin, H., Wang, M.D., Svoboda, K., Landick, R., Block, S.M., and Gelles, J. (1995). Transcription against an applied force. *Science* **270**, 1653–1657.
- Zhou, R.B., Schlierf, M., and Ha, T. (2010). Force fluorescence spectroscopy at the single-molecule level. *Methods Enzymol.* **475**, 405–426.
- Zhou, R., Kozlov, A.G., Roy, R., Zhang, J., Korolev, S., Lohman, T.M., and Ha, T. (2011). SSB functions as a sliding platform that migrates on DNA via reptation. *Cell* **146**, 222–232.



# Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo

## Graphical Abstract



## Authors

Maria Aurelia Ricci, Carlo Manzo, ..., Melike Lakadamyali, Maria Pia Cosma

## Correspondence

melike.lakadamyali@icfo.es (M.L.), pia.cosma@crg.es (M.P.C.)

## In Brief

Nucleosomes associate in discrete clutches along the chromatin fiber and clutch size correlates with cell pluripotency.

## Highlights

- Nucleosomes are arranged in heterogeneous clutches along the chromatin fiber
- The median number of nucleosomes per clutch in a given nucleus is cell-specific
- Larger and denser clutches form the “closed” heterochromatin
- Nucleosome-depleted regions separate nucleosome clutches



# Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo

Maria Aurelia Ricci,<sup>1,2,5</sup> Carlo Manzo,<sup>3,5</sup> María Filomena García-Parajo,<sup>3,4</sup> Melike Lakadamyali,<sup>3,6,\*</sup> and Maria Pia Cosma<sup>1,2,4,6,\*</sup>

<sup>1</sup>Centre for Genomic Regulation (CRG), Dr Aiguader 88, 08003 Barcelona, Spain

<sup>2</sup>Universitat Pompeu Fabra (UPF), Dr Aiguader 88, 08003 Barcelona, Spain

<sup>3</sup>ICFO, Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels, Barcelona, Spain

<sup>4</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

<sup>5</sup>Co-first author

<sup>6</sup>Co-senior author

\*Correspondence: [melike.lakadamyali@icfo.es](mailto:melike.lakadamyali@icfo.es) (M.L.), [pia.cosma@crg.es](mailto:pia.cosma@crg.es) (M.P.C.)

<http://dx.doi.org/10.1016/j.cell.2015.01.054>

## SUMMARY

Nucleosomes help structure chromosomes by compacting DNA into fibers. To gain insight into how nucleosomes are arranged in vivo, we combined quantitative super-resolution nanoscopy with computer simulations to visualize and count nucleosomes along the chromatin fiber in single nuclei. Nucleosomes assembled in heterogeneous groups of varying sizes, here termed “clutches,” and these were interspersed with nucleosome-depleted regions. The median number of nucleosomes inside clutches and their compaction defined as nucleosome density were cell-type-specific. Ground-state pluripotent stem cells had, on average, less dense clutches containing fewer nucleosomes and clutch size strongly correlated with the pluripotency potential of induced pluripotent stem cells. RNA polymerase II preferentially associated with the smallest clutches while linker histone H1 and heterochromatin were enriched in the largest ones. Our results reveal how the chromatin fiber is formed at nanoscale level and link chromatin fiber architecture to stem cell state.

## INTRODUCTION

Eukaryotic nucleosomes are a repeating unit of the chromatin, formed by 146 base pairs (bp) of DNA wrapped around octamers of the four core histone proteins (H2A, H2B, H3, and H4) (Luger et al., 1997). The histone H1 binds DNA entry/exit points of nucleosomes and to linker DNA between nucleosomes to compact the chromatin (Woodcock et al., 2006). According to the “textbook picture,” chromatin compaction follows a hierarchical model where nucleosomes form a “beads-on-string” fiber of 10 nm in diameter, which folds into higher ordered fibers of 30 nm, which in turn compact progressively into larger fibers of 100–200 nm (Finch and Klug, 1976; Song et al., 2014; Widom, 1992).

The existence of this hierarchical organization inside intact eukaryotic nuclei in vivo has recently been debated after cryo-electron microscopy, small-angle X-ray scattering (SAXS), and electron spectroscopic imaging experiments failed to detect the 30-nm fiber (Efroni et al., 2008; Fussner et al., 2012; Joti et al., 2012; Nishino et al., 2012). These studies led to the overall conclusion that the eukaryotic nuclei are mainly composed of 10 nm fibers even though the core histone proteins could not be identified unequivocally using these methods due to their lack of molecular specificity. In addition, genome-wide analyses have revealed that nucleosomes are depleted at promoter and terminator regions and at many enhancers (Struhl and Segal, 2013). Since the 30-nm fiber arrangement imposes specific constraints on nucleosome occupancy and positioning (Fussner et al., 2011a), genome-wide analyses along with the latest imaging results argue against a hierarchical organization of nucleosomes along the chromatin fiber. However, due to the limitations of previous approaches, which either lack molecular specificity or are based on population studies, histones have not been specifically visualized in intact nuclei and thus the organization of nucleosomes along the chromatin fiber has not been resolved so far.

Here, we used super-resolution nanoscopy (stochastic optical reconstruction microscopy [STORM]) (Rust et al., 2006) to visualize the structure of the chromatin fiber of a large variety of different cells at single cell level with a resolution of ~20 nm by imaging the core histone protein H2B. Super-resolution has previously been used to visualize chromatin in interphase (Bohn et al., 2010; Wombacher et al., 2010) and in dividing nuclei (Matsuda et al., 2010). Up to date, however, super-resolution studies of chromatin have not addressed questions regarding the organization of single or groups of nucleosomes, the overall nucleosome occupancy level of DNA and whether these parameters are consistent with the 30-nm fiber. Moreover, how the chromatin organization changes at the nanoscale level as a function of cell state such as pluripotent or differentiated state, while of fundamental significance for DNA accessibility and gene expression, has not yet been addressed. Overall, a quantitative approach that can estimate the number of nucleosomes within the chromatin fiber and thus identify nucleosome spatial arrangement has been lacking.

Our observations indicate that nucleosomes are grouped in discrete domains along the chromatin fiber, which we termed “nucleosome clutches” in analogy with “egg clutches.” Clutches are interspersed with nucleosome-depleted regions and the number of nucleosomes per clutch is very heterogeneous in a given nucleus arguing against the existence of a well-organized and ordered fiber. These observations were validated by computer simulations, which were also used to estimate the nucleosome occupancy of the chromatin fiber. Two-color STORM showed increased levels of H1 in larger and denser clutches containing more nucleosomes, which formed the “closed” heterochromatin. On the other hand, “open” chromatin was formed by smaller and less dense clutches which associated with RNA Polymerase II. Strikingly, despite the heterogeneity in clutch size in a given nucleus, on average differentiated cells contained larger and denser clutches compared to stem cells. These results reveal the nanoscale architecture of the chromatin fiber by showing how nucleosomes are arrayed in intact interphase nuclei.

## RESULTS

### Nucleosomes in Interphase Nuclei of Human Somatic Cells Are Organized in Discrete Nanodomains

To reveal the organization of chromatin at nanoscale resolution, we recorded STORM images of the core histone protein H2B in interphase human fibroblast nuclei (hFb) since H2B is one of the histones with fewer tail modifications and functional variants with known function (Kamakaka and Biggins, 2005). STORM images revealed a striking organization of H2B inside the nucleus (Figure 1A, left), which was not evident with conventional fluorescence microscopy (Figure S1A). H2B appeared clustered in discrete and spatially separated nanodomains (Figure 1A, left zooms). The H2B nanodomain density (number of nanodomains per unit area) was ~25% higher in the nuclear periphery, where the heterochromatin is thought to be located, compared to the nuclear interior. Since H2B is a core histone of the nucleosome octamer, its localization should reflect the arrangement of nucleosomes within the chromatin fiber. Accordingly, another core histone protein of the nucleosome octamer, H3, was similarly clustered in discrete nanodomains (Figure S1B). Furthermore, as expected, ~85% of H3 co-localized with H2B (Figure S1C).

To rule out the possibility that the observed clustered distribution of H2B was due to sample preparation or labeling methods used, we performed a series of control experiments. First, the clustered distribution of H2B was independent of the fixation and permeabilization protocols used (Figures S1D and S1E). Second, STORM images contained discrete nanodomains when H2B was indirectly labeled using an antibody against SNAP tag in cells stably expressing H2B-SNAP (Figure S1F). Third, we ruled out potential artifacts in H2B STORM images associated with the large size of the antibody by comparing to nanobody labeling (Figures S1G–S1N). Fourth, labeling efficiency defects were also ruled out by computer simulations of nucleosome arrangements (see further details in [Extended Experimental Procedures](#) and [DNA Fiber Is Not Fully Occupied with Nucleosomes](#) section). Finally, to confirm the existence of H2B nanodomains in living cells we imaged H2B-mEos2 or

H2B-PA-mCherry expressing hFbs. In both cases super-resolution imaging revealed discrete and spatially separated nanodomains as in the case of fixed cells (Figures 1B and S1O).

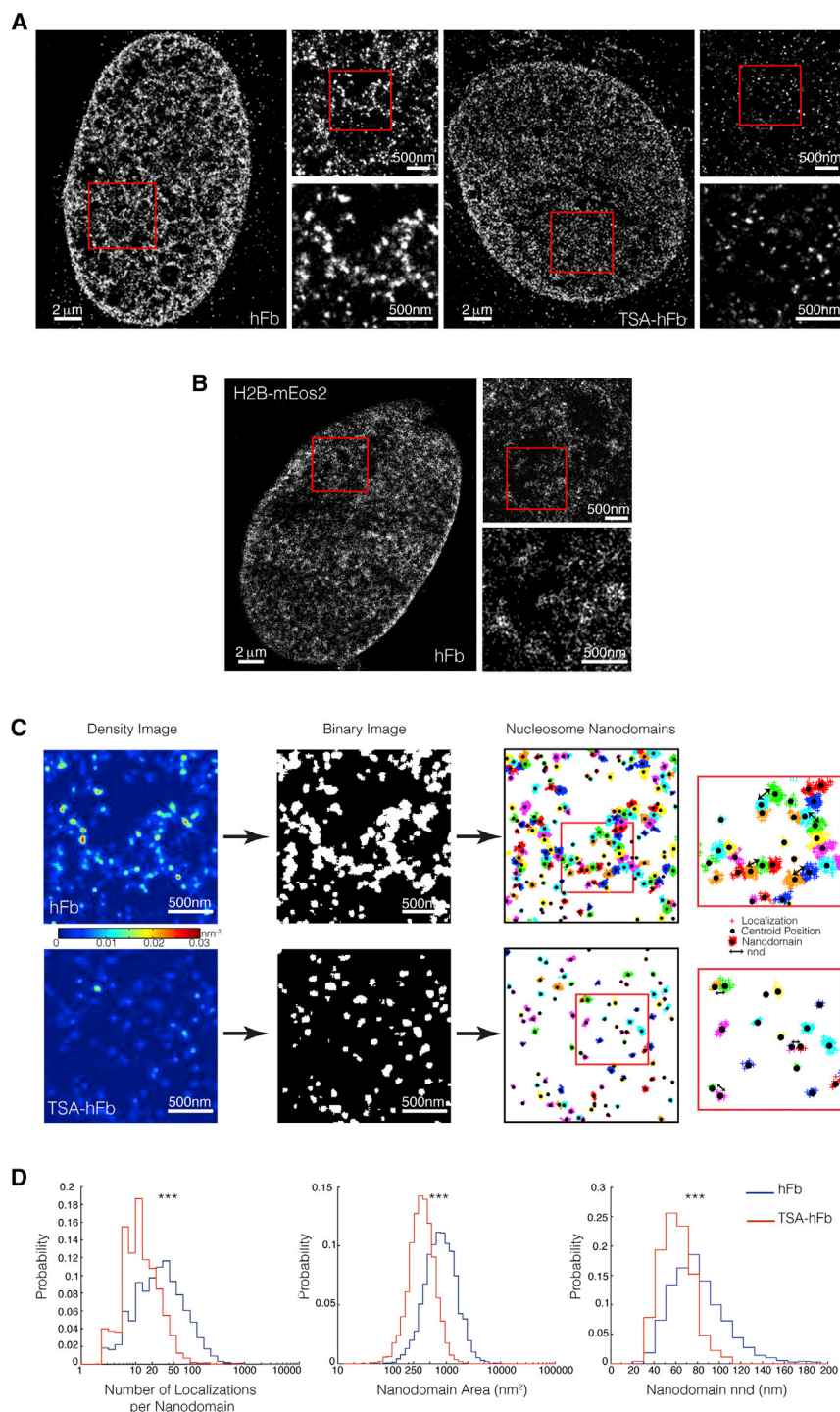
We next analyzed the nucleosome organization in cells undergoing massive epigenome modifications and chromatin rearrangements. For this, hFbs were treated with Trichostatin A (TSA) (TSA-hFb), a potent inhibitor of histone deacetylase enzyme, which leads to genome-wide decondensation of chromatin through accumulation of acetylation groups on histone tails (Tóth et al., 2004). As expected, there was a large increase in H3 acetylation after TSA treatment (Figure S1P). TSA treatment also resulted in visually evident changes in the nuclear distribution of H2B nanodomains (Figure 1A, right), which appeared dimmer and hence contained less localizations. Furthermore, the nanodomains were also more dispersed within the nucleus (Figure 1A, right zooms). The H2B nanodomain density was enhanced by ~10% in the nuclear periphery of TSA-hFbs compared to the nuclear interior, although it was less dense than the nuclear periphery of untreated hFbs. Finally, the distribution of acetylated H3 was also highly dispersed in the nuclei, mirroring the spatial re-distribution observed for the H2B nanodomains after TSA treatment (Figure S1P). These changes overall indicate that nucleosomes undergo spatial rearrangement in hFb nuclei upon chromatin decondensation.

To gain quantitative insight into the H2B nanodomains, we next developed a cluster identification algorithm to group the localizations in STORM images into nanodomains ([Extended Experimental Procedures](#); Figures 1C and S1Q). Quantitative analysis revealed that the distributions of the number of localizations per nanodomain, nanodomain areas, and nanodomain nearest neighbor distances (nnDs) were shifted to lower values in TSA-hFbs compared to hFbs (Figure 1D), and hence nucleosomes showed statistically significant spatial re-organization after TSA treatment and chromatin decondensation.

In control experiments, nanodomain areas of hFbs were similar when H2B was labeled with an antibody (Mean Area  $\pm$  SEM =  $830 \pm 70$  nm<sup>2</sup>,  $n = 11$  cells), with GFP-nanobody in hFbs transfected with H2B-GFP (Mean Area  $\pm$  SEM =  $660 \pm 70$  nm<sup>2</sup>,  $n = 7$  cells,  $p = 0.1760$ ) and in living or fixed hFbs expressing H2B-mEos2 or H2B-PA-mCherry (Mean Area  $\pm$  SEM =  $660 \pm 30$  nm<sup>2</sup> in living cells,  $n = 12$  cells,  $p = 0.068$  and  $610 \pm 40$  nm<sup>2</sup> in fixed cells,  $n = 5$  cells,  $p = 0.1088$ , Figure S1R), indicating that the large size of the antibody or fixation did not significantly affect the spatial resolution of H2B STORM images or the organization of nanodomains. The number of localizations per nanodomain was lower when using fluorescent proteins compared to organic fluorophores as expected (Figure S1R), since mEos2 and PA-mCherry are known to undergo less blinking and photoactivate with only moderate efficiency (Durisic et al., 2014) compared to AlexaFluor647.

### Wild-Type Mouse Embryonic Stem Cells Cultured under Different Media Conditions and Mutants Have Distinct Nucleosome Organization in Interphase

To assess the nucleosome organization of pluripotent cells, we next imaged H2B in mouse embryonic stem cells (mESCs). mESCs were cultured under two different media conditions: (1) with serum and the cytokine leukemia inhibitory factor (sLif),



**Figure 1. Nucleosomes Are Arranged in Discrete Nanodomains in Interphase Nuclei of Human Somatic Cells**

(A) Representative STORM images of H2B in human fibroblast nucleus (hFb, left) ( $n = 11$  cells) and Trichostatin A (TSA)-treated human fibroblast nucleus (TSA-hFb, right) ( $n = 11$  cells). Progressively higher zooms of the regions inside the red squares are shown next to each nucleus.

(B) Live cell super-resolution images of hFbs expressing H2B-mEos2. Progressively higher zooms of the regions inside the red squares are shown next to each nucleus.

(C) Density images showing regions of high (red) and low (blue) H2B density (number of H2B localizations per unit area) in hFb (upper) and TSA-hFb (lower) according to the color scale bar. After thresholding, the density images are converted into binary images in which regions containing H2B localizations appear white. Every white region is analyzed using a cluster identification algorithm that groups the individual localizations based on their proximity into nanodomains. Shown are example nanodomains in hFb (upper) and TSA-hFb (lower) for which localizations (crosses) having the same color belong to the same nanodomain. The centroid position of each nanodomain is shown as a black dot. The nearest neighbor distances (nnds) between nanodomains inside the white regions are calculated (double head black arrows), along with the number of localizations per nanodomain and the nanodomain area.

(D) Representative distributions of the number of H2B localizations per nanodomain, nanodomain area, and nnds between nanodomains in hFb (blue) and TSA-hFb (red) for the cells shown in (A). Statistical significance between the different distributions is shown as \*\*\* ( $p < 10^{-3}$ ).

See also Figure S1.

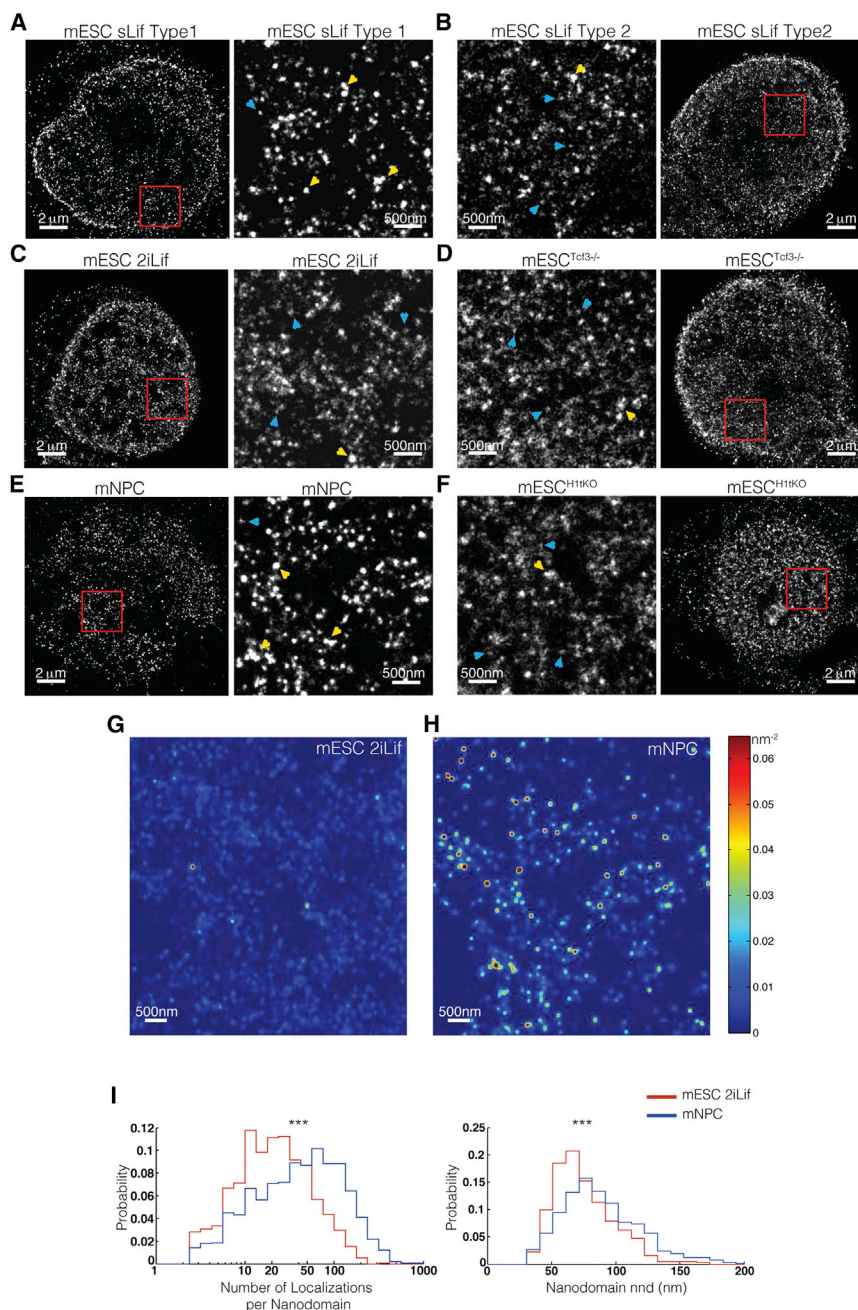
gram of their transcriptional profile and a more homogenous expression of pluripotency factors (Marks et al., 2012; Wray et al., 2010).

As expected, mESCs cultured in sLif expressed varying levels of the pluripotency marker Nanog (Figure S2A). Low Nanog expressing cells (Figure S2A, upper) had bright nanodomains in STORM images (i.e., containing a large number of localizations) (Figure 2A, type 1, yellow arrowheads). On the other hand, high Nanog expressing mESCs cultured in sLif (Figure S2A, lower) mostly had dim

and (2) with inhibitors of two kinases (Mek and Gsk3) known as “2i” and Lif (2iLif). mESCs cultured in sLif have heterogeneous morphology, exhibit heterogeneous expression of pluripotency factors (Cahan and Daley, 2013), and display appreciable expression of ectoderm and mesoderm genes (Marks et al., 2012). On the other hand, 2iLif maintains mESCs in a ground-state (Ying et al., 2008), characterized by no predetermined pro-

nanodomains (Figure 2B, type 2, cyan arrowheads). In addition, the nanodomains appeared more dispersed inside the nucleus. Nanodomains of mESCs cultured in 2iLif and of mESCs<sup>Tcf3<sup>-/-</sup></sup> were mostly dim (Figures 2C, 2D, and 2G). Similar to 2iLif, the deletion of Tcf3 (mESCs<sup>Tcf3<sup>-/-</sup></sup>), a key effector of the Wnt/ $\beta$ -catenin pathway, was also previously shown to maintain the ground-state of pluripotency (Cole et al., 2008; Tam et al.,





**Figure 2. Nucleosomes Are Arranged in Discrete Nanodomains in Interphase Nuclei of Mouse Embryonic Stem Cells**

(A–F) Representative STORM images of H2B in (A) type 1 mouse embryonic stem cells (mESCs) cultured in serum plus Lif (sLif) ( $n = 8$  cells), (B) type 2 mESCs cultured in sLif ( $n = 6$  cells), (C) mESCs cultured in 2iLif ( $n = 15$  cells), (D) mutant mESCs lacking Tcf3 (mESC<sup>Tcf3-/-</sup>) ( $n = 10$  cells), (E) neuronal precursor cells (mNPC) obtained after differentiation of mESCs ( $n = 9$  cells), and (F) mutant mESCs that are triple H1 knockout (mESC<sup>H1tkO</sup>) ( $n = 15$  cells). Next to each cell type, higher zooms of the regions inside the red squares are shown. Yellow arrowheads point to bright nanodomains comprising a large number of localizations whereas cyan arrowheads point to dimmer nanodomains comprising a small number of localizations.

(G and H) Density image showing the differences in nanodomain organization of mESCs cultured in 2iLif (G) and mNPCs (H). Regions of high (red) and low (blue) H2B density are shown according to the color scale bar.

(I) Representative distributions of the number of H2B localizations per nanodomain and nanodomain nnds in mESCs cultured in 2iLif medium (red) and mNPCs (blue) for the cells shown in (C) and (E). Statistical significance is shown as \*\*\* ( $p < 10^{-3}$ ). See also Figure S2.

Taken together, these results indicate that the chromatin in ground-state mESCs is characterized by dimmer H2B nanodomains, which are more dispersed inside the nuclear space and by increased acetylation level.

The linker histone H1 is thought to play an important role in chromatin organization and higher order compaction (Clau-sell et al., 2009; Woodcock et al., 2006). mESCs carrying a deletion of three H1 isoforms (mESC<sup>H1tkO</sup>), which were shown to have reduced chromatin compaction (Fan et al., 2005) contained a large amount of dim nanodomains (Figure 2F) having a similar organization to those observed in mESCs cultured in 2iLif and in mESCs<sup>Tcf3-/-</sup>.

Quantitative analysis also confirmed that the number of localizations per nanodomain and nanodomain nnds were lower in ground-state mESCs with respect to somatic mNPCs (Figures 2I and 3, below).

#### Nanodomains Contain a Discrete Number of Nucleosomes and the Nucleosome Number Correlates with Pluripotency

Given the identical labeling and imaging conditions used for each cell type (Extended Experimental Procedures; Table S1), the number of nucleosomes should scale with the number of

2008; Yi et al., 2008). When mESCs were differentiated into neural precursor cells (mNPCs) the H2B nanodomains became brighter, resembling those observed in hFbs (Figures 2E and 2H).

ESCs<sup>Tcf3-/-</sup> were shown to contain large epigenome modifications (Lluis et al., 2011). Accordingly, there was increased level of acetylation in these cells (Figure S2B) with respect to type 1 mESCs cultured in sLif (Figure S2C). mESCs cultured in 2iLif (Figure S2D) as well as type 2 mESCs cultured in sLif (Figure S2E) also contained higher levels of H3 acetylation, while mNPCs showed a lower level of H3 acetylation (Figure S2F).

localizations (Dani et al., 2010). Nanodomains in any given nucleus contained a large distribution of localizations spanning two orders of magnitude ( $\sim 3$  to 300) (Figures 1D and 2I), indicating that they comprised heterogeneous groups with varying numbers of nucleosomes. We will refer to these heterogeneous nucleosome groups as “nucleosome clutches” in analogy to “egg clutches” and we will use the term “clutch size” interchangeably with the number of nucleosomes per clutch. Despite this heterogeneity, the median number of localizations per clutch in individual cells correlated strongly with cell type and showed statistically significant differences between hFbs and TSA-hFbs and among the different mESCs (Figures 3A and 3B). Control experiments showed that the median number of localizations per clutch in hFbs was similar when H3 was labeled ( $N_{\text{localizations}} = 24 \pm 2$ ) instead of H2B ( $N_{\text{localizations}} = 24 \pm 4$ ) and under different fixation and permeabilization conditions ( $N_{\text{localizations}} = 24 \pm 4$  for ethanol/methanol fixation,  $N_{\text{localizations}} = 26 \pm 3$  for PFA fixation), excluding potential sample labeling artifacts.

Overall, the differences in the median numbers of localizations indicate that nucleosomes assemble into clutches of larger size in hFbs compared to TSA-hFbs (Figure 3A). Similarly, nucleosomes formed larger clutches in differentiated mNPCs and mESCs cultured in sLif compared to mESCs cultured in 2iLif, mESC<sup>Tcf3<sup>-/-</sup></sup> and mESC<sup>H1tKO</sup> (Figure 3B).

In order to relate the median number of localizations to the median number of nucleosomes in different cell types, we further generated a calibration curve by imaging in vitro-labeled mononucleosomes and polynucleosome arrays containing 12- or 24-nucleosomes (Grigoryev et al., 2009) (Extended Experimental Procedures; Figures S3A–S3C). Mononucleosomes had a median number of ten localizations, indicating a high detection efficiency of single nucleosomes using STORM. We also labeled and imaged the 12- and 24-polynucleosome arrays in the presence of nuclear extract to better emulate the crowding of the nuclear environment (Extended Experimental Procedures). A similar median number of localizations was obtained in the presence of the extract (Figures 3C, S3B, and S3D) reassuring that labeling efficiency does not significantly differ under both conditions. The calibration curve was also validated by imaging a plasmid with a length allowing the assembly of  $\sim 20$  nucleosomes. The median number of localizations obtained corresponded to  $19.5 \pm 2$  nucleosomes after interpolation, confirming that the calibration curve was indeed accurate (Figure 3C). We also estimated that on average 1.6 antibodies (1/0.6) were present on one mononucleosome (Figures 3C, inset, S3A, and S3B). We note that even when the antibody binding efficiency was similar in the absence and presence of nuclear extract, we cannot fully exclude some underestimation in the nucleosome numbers, in particular for the larger clutches. Nevertheless, this underestimation should not affect the relative comparison among the different cell types.

We next used the calibration curve to estimate the median number of nucleosomes per clutch (Figure 3D). Clutches in hFbs comprised a median of  $\sim 8$  nucleosomes whereas this number decreased to  $\sim 2$  nucleosomes after TSA treatment (Figure 3D, left). mESCs cultured in sLif constituted a heterogeneous population compared to other mESCs, consisting of cells with a median of  $\geq 4.5$  nucleosomes (type 1 mESCs, corresponding to

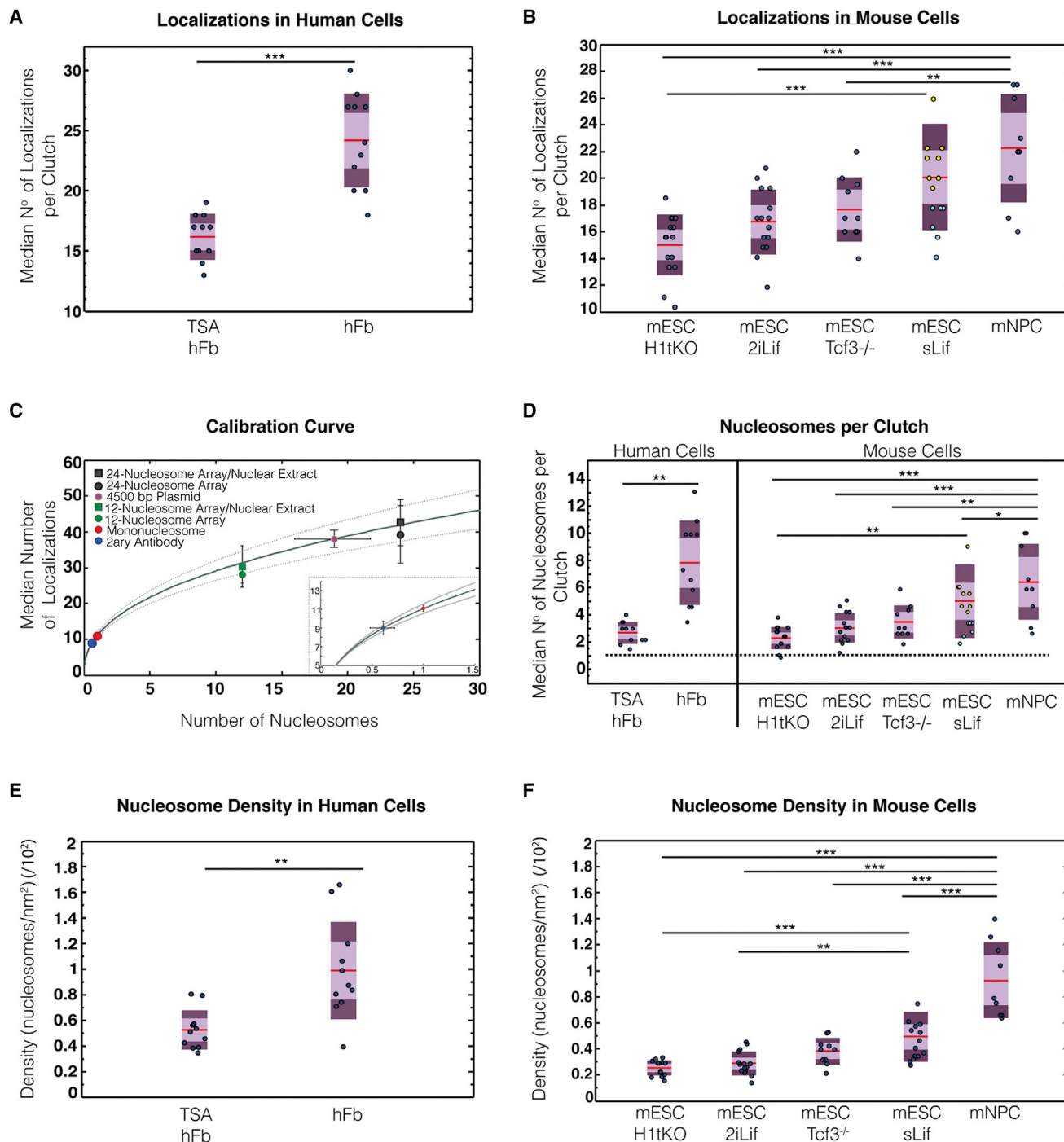
$22 \pm 2$  localizations) and cells with a median of  $< 4.5$  nucleosomes per clutch (type 2 mESCs, corresponding to  $17 \pm 2$  localizations) (Figure 3D, right; Extended Experimental Procedures). mNPCs were also heterogeneous and had clutches with on average a larger number of nucleosomes ( $\sim 6$ , Figure 3D, right). The number of nucleosomes per clutch was less variable in mESCs cultured in 2iLif, mESC<sup>Tcf3<sup>-/-</sup></sup>, and mESC<sup>H1tKO</sup> (median of  $\sim 3$ ,  $\sim 3.5$ , and  $\sim 2$ , respectively) (Figure 3D, right). These results indicate that nucleosomes are assembled together in smaller clutches in pluripotent cells and in increasing numbers in differentiated cells. Furthermore, clutch size drastically changes upon chromatin decondensation after TSA treatment.

hFbs had more densely compacted nucleosome clutches compared to TSA-hFbs (Figure 3E) as determined from the median nucleosome density (number of nucleosomes per unit area). Nucleosome density was likewise higher for mNPCs and mESCs cultured in sLif with respect to mESCs cultured in 2iLif, mESC<sup>Tcf3<sup>-/-</sup></sup> and mESC<sup>H1tKO</sup> (Figure 3F). Therefore, nucleosome density is in general low in pluripotent cells and nucleosome compaction increases upon differentiation.

### Clutch Size Correlates with the Pluripotency Grade of Human-Induced Pluripotent Stem Cells

Next, we aimed to study whether the number of nucleosomes per clutch could be predictive of the pluripotency grade in human-induced pluripotent stem cell (hiPSCs) clones, as defined by their gene expression profile and propensity to differentiate. hiPSCs were generated from hFbs and characterized using standard methods (Figures S4A–S4D). The hiPSC clone 13 and 8 were both pluripotent since they were AP-positive and expressed the stem cell markers TRA1-60, SSEA4, Oct4, Sox2, and Nanog. However, while the hiPSC clone 13 formed embryoid bodies, which differentiated into the three germ layers, and generated large and fully differentiated teratomas in mice, the hiPSC clone 8 did not form the ectoderm layer from the embryoid bodies and it generated very small undifferentiated teratomas in vivo (Figures S4A–S4D). Furthermore, the Oct4 expression level of single cells in the hiPSC clone 8 was 14-fold lower compared to hiPSC clone 13 (Figure S4B). Therefore, the pluripotency grade of clone 13 was higher compared to clone 8. To rank the pluripotency grade of all hiPSC clones in a more quantitative manner, we used the gene card technology that gives a pluripotency score based on expression level of stemness genes and differentiation propensity compared to a reference set of formerly characterized human embryonic stem cell (hESC) and hiPSC lines (Bock et al., 2011). The gene card results agreed with the classical characterization of clones 8 and 13 and allowed quantitative ranking of the remaining hiPSC clones in order of pluripotency grade (Figure S4E).

The median number of localizations quantified from STORM images (Figure 4A) showed statistically significant differences among the different clones and gradually increased passing from the hiPSCs clone 13 to 8. The calibration curve was used to deduce the median number and density of nucleosomes inside clutches in each hiPSC clone (Figures 4B and 4C). There was a remarkable agreement between the pluripotency score obtained from the gene card and the clutch size

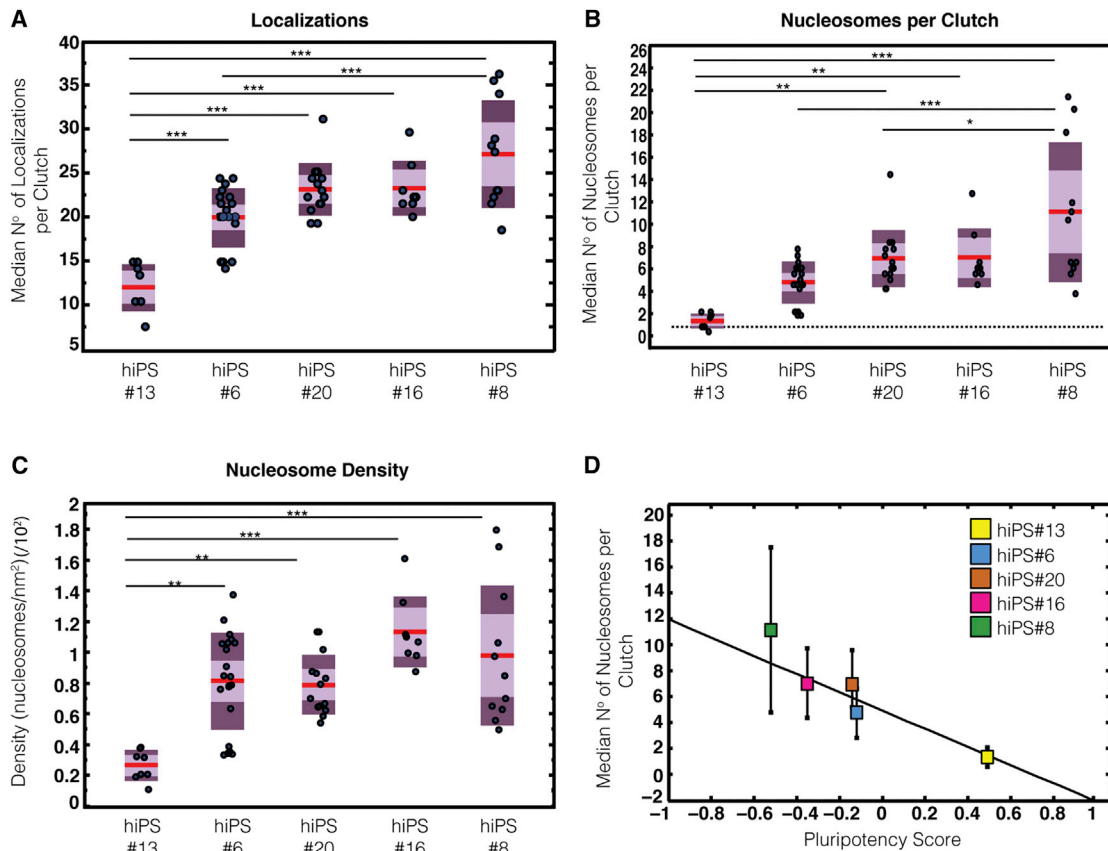


**Figure 3. The Number of Nucleosomes Inside Clutches Correlates with Cellular State**

(A and B) Box plots showing the median number of H2B localizations per clutch in hFbs ( $n = 11$  cells), TSA-hFbs ( $n = 11$  cells) (A), in different mESCs ( $n = 15, 15, 10$ , and 14 cells, respectively, from left to right), and in mNPCs ( $n = 9$  cells). (B) mESC sLif cells are color coded as type 1 containing a median of  $22 \pm 2$  localizations (yellow) ( $n = 8$  cells) and type 2 containing a median of  $17 \pm 2$  localizations (cyan) ( $n = 6$  cells).

(C) Calibration curve to deduce the median number of nucleosomes per clutch. The median number of localizations per mononucleosome (red circle), 12- (green circle) and 24-nucleosome array (black circle) labeled and imaged in vitro, 12- (green square) and 24-nucleosome array (black square) labeled and imaged in the presence of a nuclear extract were used to generate the calibration curve. The gray line is the fit for the data in the presence of nuclear extract to a power law  $y = ax^b$  with  $a = 11 \pm 3$  and  $b = 0.41 \pm 0.15$ . Errors correspond to 95% confidence bounds. The dotted lines represent 68% confidence interval. Purple circle is data from a 4,500 base pair (bp) plasmid assembled into nucleosome-arrays with an expected number of  $\sim 20$  nucleosomes per array. Blue circle is data from

(legend continued on next page)



**Figure 4. Clutch Size Correlates with Pluripotency Grade in Human-Induced Pluripotent Stem Cells Clones**

(A) Box plots showing the median number of H2B localizations per clutch in different human-induced pluripotent stem cell (hiPSCs) clones (n = 8, 20, 14, 8, and 11 cells from left to right, respectively, from multiple imaging experiments [minimum of 3]).

(B) Box plots showing the median number of nucleosomes per clutch in the different hiPSCs. The dotted line corresponds to one nucleosome.

(C) Box plots showing the median density of nucleosomes per clutch in the different hiPSCs.

(D) Pluripotency score of the different hiPSCs obtained from the gene card plotted against the median number of nucleosomes. Error bars indicate SDs. For black dots, lines, box plot colors and statistics in (A)–(C) see description in the legend of Figure 3.

See also Figure S4.

(Figure 4D), (analysis showed  $r = -0.94$  indicating high level of anticorrelation, i.e., low number of nucleosomes per clutch for high pluripotency score and vice versa). Indeed, the hiPSC clone 13, which showed high propensity to differentiate and a high pluripotency score, had low density clutches with a median number of only 1 nucleosome, while clutch size and density increased progressively with the decreased pluripotency score (Figures 4B–4D).

### Larger Clutches Have Higher Levels of H1 and Lower Levels of RNA Polymerase II

The arrangement of nucleosomes in small clutches with lower compaction could potentially facilitate the binding of transcription factors, polymerases, and other proteins to the DNA, which should be more accessible in regions containing smaller clutches. The higher compaction of the nucleosomes within larger clutches, on the other hand, should restrict DNA

fluorophore-labeled secondary antibody alone. Inset shows the first part of the curve containing the secondary antibody and the mononucleosomes. Error bars correspond to SDs.

(D) Box plots showing the median number of nucleosomes per clutch in hFbs, TSA-hFbs, in the different types of mESCs and mNPCs. The dotted line corresponds to one nucleosome.

(E and F) Box plots showing the median density of nucleosomes per clutch in hFbs, TSA-hFbs (E) in the different types of mESCs and mNPCs (F). For (A), (B), and (D–F) each black dot shows the median number of nucleosomes obtained per individual nucleus from multiple imaging experiments (minimum of 3). The red line is the median for the entire population of nuclei analyzed for that cell type. The light magenta region corresponds to the SE and the dark magenta region to the SD. Statistical significance between the different cell types was determined using one-way ANOVA. The stars indicate p values according to \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), and \*\*\* ( $p < 0.001$ ).

See also Figure S3.



accessibility and should be aided by the presence of linker histone protein H1, which is known to be involved in nucleosome compaction and is enriched in heterochromatin (Fan et al., 2005; Woodcock et al., 2006). Thus, to evaluate differences in the heterochromatin content of clutches and their accessibility to RNA Polymerase II (PolII), we performed multi-color STORM imaging of H2B with histone H1 and of H2B with PolII.

H1 was more enriched at the nuclear periphery of hFbs where heterochromatin is more abundant (Meister and Taddei, 2013) (Figure 5A). A higher percentage of H2B co-localized with H1 in hFbs ( $61\% \pm 11\%$ ) compared to TSA-hFbs ( $42\% \pm 6\%$ ) ( $p = 0.028$ ) as is also evident in the zoomed images (Figures 5A and 5B). For both hFbs and TSA-hFbs, the number of H1 localizations in the clutches increased with the number of H2B localizations (Figures 5C and S5A). In mESCs cultured in sLif,  $\sim 54\% \pm 2\%$  of H2B co-localized with H1 and the number of H1 localizations also increased with the number of H2B localizations (Figure S5B). As expected, mESCs<sup>H1<sup>1KO</sup></sup> contained much lower amount of H1 (Figure S5C) and only  $\sim 35\% \pm 4\%$  of H2B co-localized with H1 ( $p = 0.0057$ ). Despite the low amount of H1 in these cells, the same trend was observed, i.e., the number of H1 localizations was increased in clutches with an increasing number of H2B localizations (Figure S5B). These results overall suggest that the number of H1 histones correlates with the number of nucleosomes inside the clutches.

Since the largest clutches containing high amounts of H1 were also the more densely compacted ones (Figures 3E and 3F) we hypothesized that these might correspond to the ‘closed’ heterochromatin regions. To test this hypothesis we used an anti-CREST antibody to recognize specific centromeric proteins. Centromeres are known to include heterochromatin (Meister and Taddei, 2013). CREST positive regions co-localized with the large clutches (Figure 5D) containing on average 1.3-fold higher number of H2B localizations compared to the global median ( $p = 0.014$ ) (Figure S5D). A similar analysis was performed in mESCs expressing a TALE-mClover that accumulates at pericentromeric regions in these cells (Miyanari et al., 2013). mClover positive regions once again correlated with large clutches (Figure 5E) and clutches that overlapped with TALE-mClover contained on average 2.2-fold higher number of H2B localizations compared to the global median ( $p = 0.0002$ ) (Figure S5E).

Next we analyzed PolII and H2B multi-color STORM images of hFbs and TSA-hFbs. In both cases, PolII was partially interspersed and partially co-localized with the nucleosome clutches (Figure 6A and zooms). PolII-H2B nnds peaked at  $\sim 40$  nm (Figure 6B). We rationalized that the DNA within clutches having fewer nucleosomes should be more accessible and therefore PolII should be closest to the small clutches. To test this hypothesis, we analyzed the number of H2B localizations within clutches as a function of the nnds between PolII and H2B, restricting the analysis to nnds below 70 nm, which corresponds to the maximum PolII cluster size plus the maximum clutch size. For both hFbs and TSA-hFbs, the nnds between PolII and H2B were shorter for smaller clutches, indicating that PolII was indeed closer to the smaller clutches with few nucleosomes (Figure 6C). These results indicate that PolII can access small clutches, which likely form the “open” chromatin fiber arrangement of transcribed chromatin regions.

### The DNA Fiber Is Not Fully Occupied with Nucleosomes

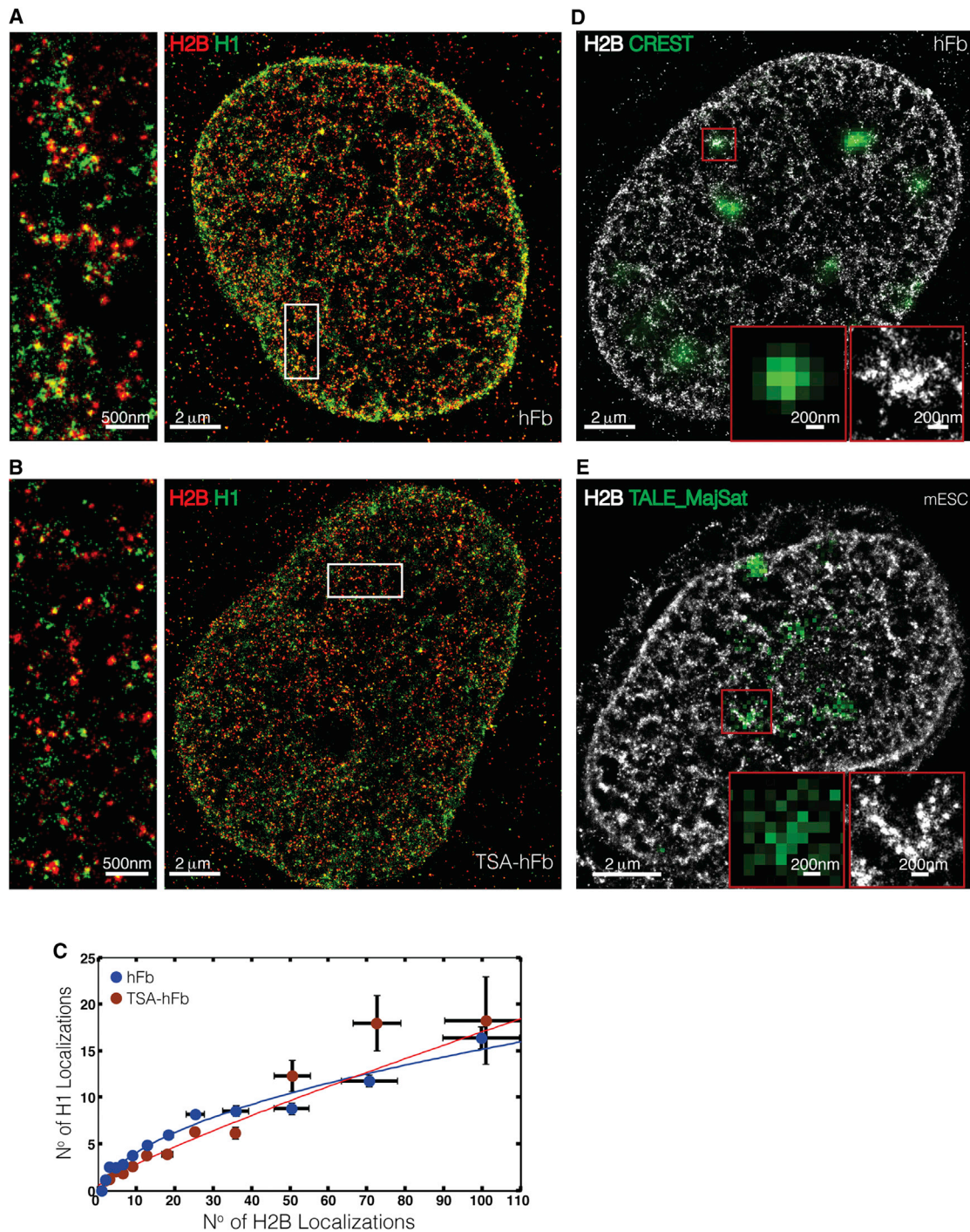
The organization of nucleosomes in discrete, spatially separated clutches implies that nucleosome-depleted regions likely exist in the chromatin fiber. We hypothesized that these regions might be due to removal of nucleosomes in between nucleosome-rich regions or to variations in the length of the linker-DNA between subsequent nucleosomes. Coarse-grained computer simulations of nucleosome spatial arrangement were performed, using a simplistic model that considers a minimum number of parameters (Extended Experimental Procedures; Figures 7A–7C). In this model, we simulated either random removal of nucleosomes with a given probability (NR Model; Extended Experimental Procedures; Figure 7D) or variations in the average length of the linker-DNA (LL Model, Extended Experimental Procedures; Figure 7E) or potential effects of incomplete labeling (Extended Experimental Procedures; Figures S6A and S6B).

Synthetic STORM images of the nucleosomes along the DNA fiber (Figures 7F and S6A) were generated by assigning to each nucleosome a given number of localizations based on the in vitro calibration results (Extended Experimental Procedures; Figure S3B). The synthetic STORM images at different nucleosome occupancy levels (Figure 7F) showed striking resemblance to the experimental images. The median number of localizations, area, and nnds of the nucleosome clutches were determined using identical analysis parameters as before and plotted as a function of nucleosome occupancy (Figure 7G).

Both the NR and LL models intersected the experimental values of the number of localizations and the clutch nnds at  $\sim 57\%$  and  $\sim 45\%$  occupancy for the hFbs and TSA-hFbs, respectively (Figure 7G, top and middle). For TSA-hFb, the NR model intersected the experimental value of the clutch area at a similar occupancy level (45%) whereas the LL model intersected it at a much lower occupancy level (34%) (Figure 7G, bottom). For hFbs, the NR model intersected the experimental value of clutch area at a slightly higher occupancy level than those obtained from the other two parameters (60%) whereas the LL model intersected this value at a slightly lower occupancy level (52%) (Figure 7G, bottom).

In the case of labeling efficiency simulations, the three measured experimental parameters could not be simultaneously reproduced at any given labeling efficiency for hFbs and TSA-hFbs (Figure S6B), indicating that poor labeling efficiency alone cannot explain the experimental observations. However, nucleosome depletion in combination with incomplete labeling can lead to the observed results, shifting the nucleosome occupancy to higher values (Figure S6C). Regardless of the labeling efficiency, nucleosome occupancy was higher in hFbs compared to TSA-hFbs. The simulation results could reproduce both the median values observed for the experimental data as well as the full experimental distributions, with the best fit for the NR model corresponding to 75% labeling efficiency for both hFb (60% occupancy) and TSA-hFb (48% occupancy) (Figures S6D–S6F).

Taken altogether, these results indicate that linker length variations do not play a major role in generating nucleosome poor regions in TSA-hFbs since all three measured parameters of the experimental data could not be recapitulated with this model. In the case of hFbs, combination of nucleosome removal and



**Figure 5. The Linker Histone H1 Increases in Large Clutches and These Correlate with Heterochromatin Markers**

(A and B) Representative STORM images showing H2B (red) and H1 (green) in hFb ( $n = 4$  cells) (A) and TSA-hFb ( $n = 4$  cells) (B). Higher zooms of the regions inside white rectangles are shown next to each nucleus.

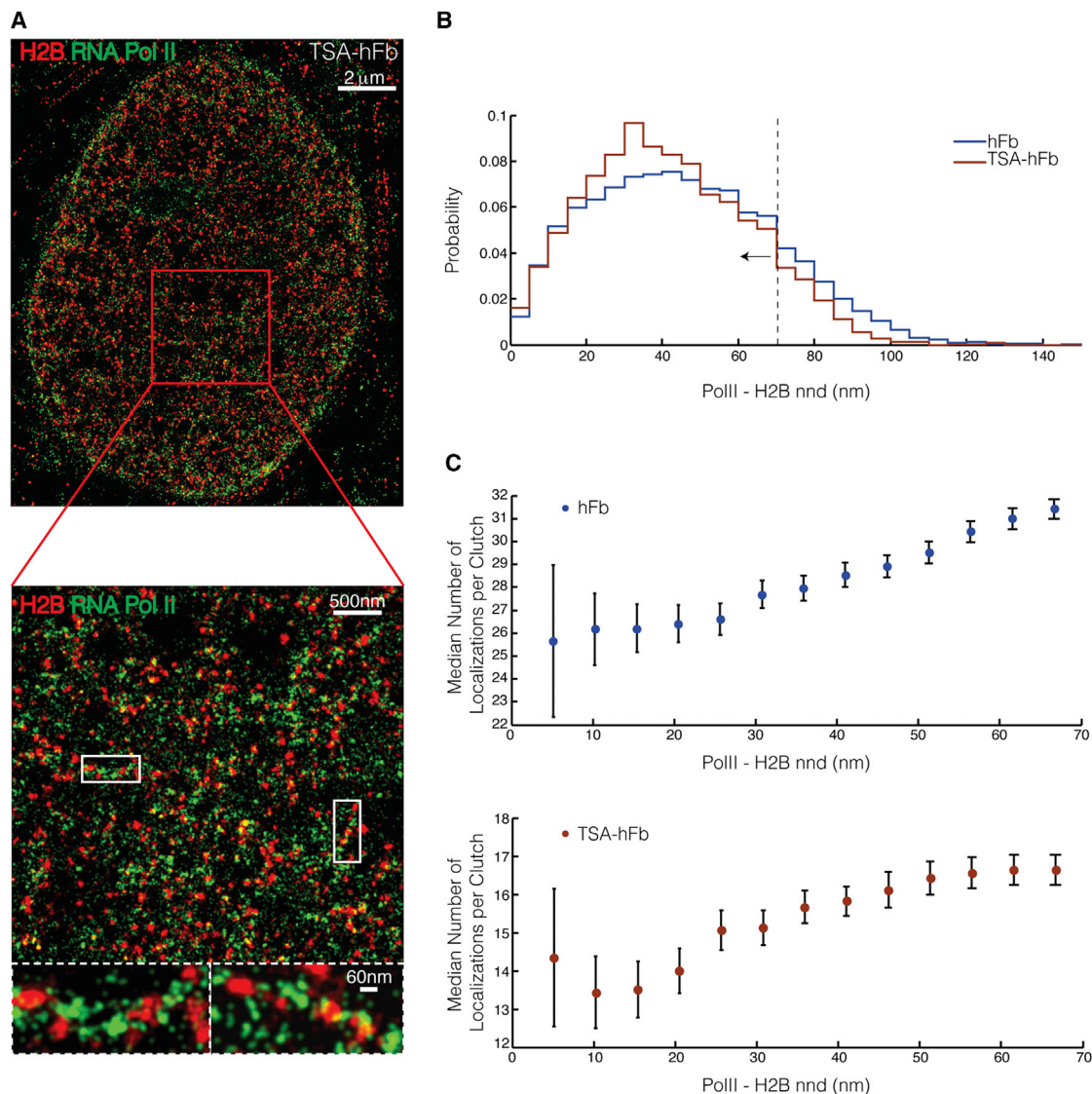
(C) Plot showing the number of H2B (x axis) and H1 (y axis) localizations inside clutches for which these two histones showed colocalization. Error bars in x axis indicate SDs and in y axis indicate SEs. The trend lines are polynomial fits intended as a guide to the eye.

(D) Representative STORM image of H2B (gray) overlaid with the conventional fluorescence image of anti-CREST antibody (green) which recognizes centromeric proteins in hFbs ( $n = 6$  cells). Inset shows a zoomed in region of the red square.

(E) Representative STORM image of H2B (gray) overlaid with the conventional fluorescence image of TALE-mClover that recognizes the major satellite of pericentromeric regions (TALE\_MajSat) (green) in mESC sLif ( $n = 16$  cells). Inset shows a zoomed in region of the red square.

See also Figure S5.





**Figure 6. RNA Polymerase II Associates with the Small Clutches**

(A) Representative STORM image showing H2B (red) and RNA polymerase II (PolII) (green) in TSA-hFb. Progressive zooms of the regions inside white rectangles are shown below the image of the nucleus.

(B) Plot showing the distribution of nnds between H2B and PolII in hFb (blue) ( $n = 5$  cells) and TSA-hFb (red) ( $n = 3$  cells). The dashed line at 70 nm shows the distance cut-off used for the analysis in (C) corresponding to maximum clutch size plus maximum PolII cluster size.

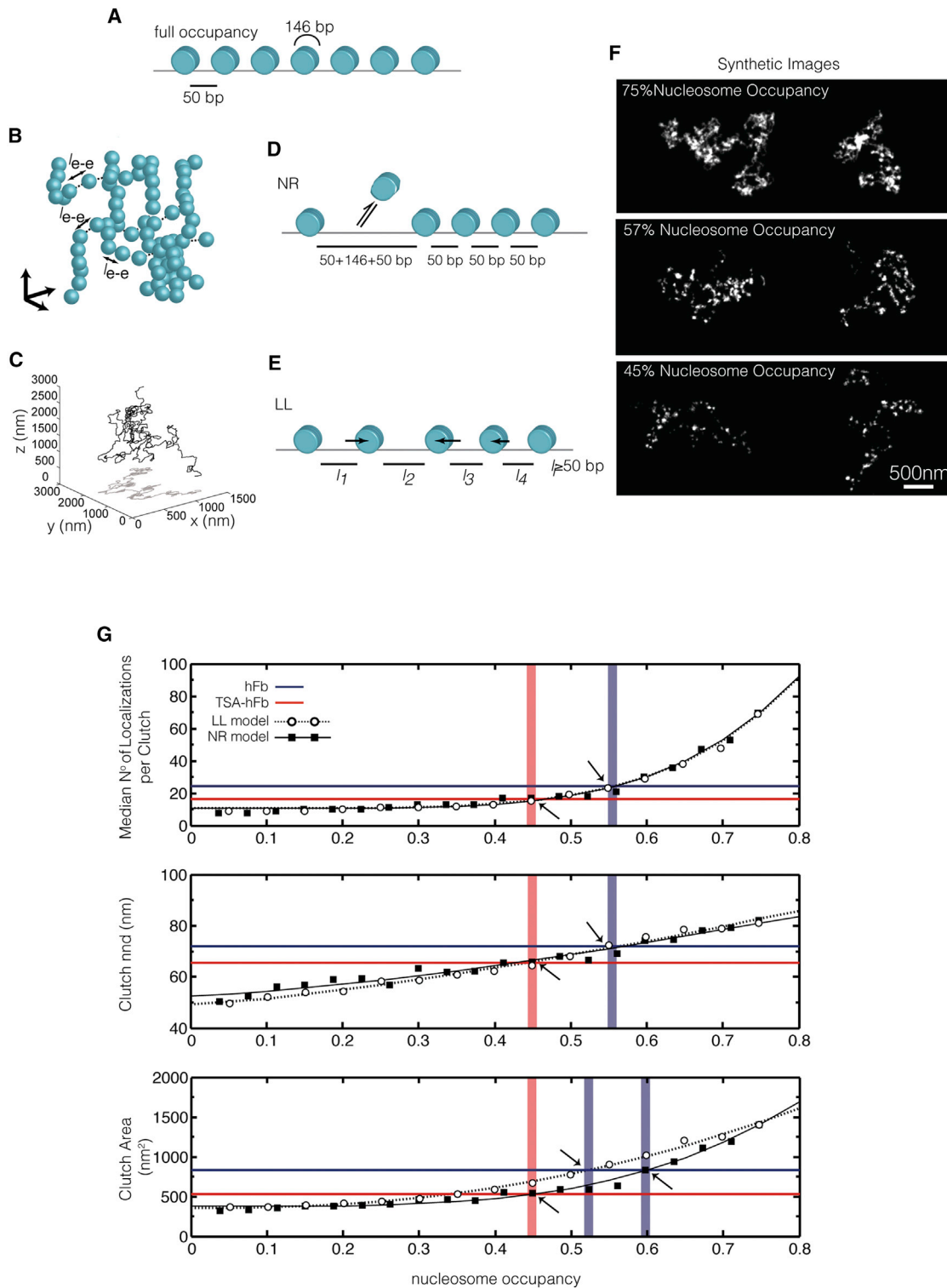
(C) Plot showing the median number of H2B localizations within clutches as a function of the nnds (up to a maximum nnd of 70 nm) between PolII and H2B for hFb (blue) and TSA-hFb (red). Error bars indicate SEs.

linker-DNA length modifications likely plays a role in generating the nucleosome-depleted regions.

## DISCUSSION

Chromatin organization and structure in interphase nuclei is important for gene function and activity, therefore it is an area of intense investigation (Wendt and Grosfeld, 2014). Electron microscopy (EM) and more recently cryo-EM (Song et al., 2014) have provided invaluable insight into nucleosome organization in vitro. However, in vitro studies cannot determine if the organi-

zation observed is prevalent in vivo in intact nuclei. The structure of chromatin has also been subject to a number of in vivo studies (Fussner et al., 2011a). Although these previous methods have provided key information, they are accompanied with major drawbacks such as harsh sample preparation, lack of molecular specificity and/or low resolution, such that a clear picture on the organization of nucleosomes along the chromatin fiber in living cells has been lacking so far. Here, we have come closer than ever to visualize the native structure of the chromatin fiber by dissecting at nanoscale resolution the organization of nucleosomes in intact nuclei and in single cells. STORM imaging revealed that



**Figure 7. Computer Simulations of Nucleosome Occupancy**

(A) Nucleosomes (light blue) are initially arranged at regular intervals of 50 bp (experimentally determined linker-DNA length) on the DNA fiber, (full occupancy, which in reality corresponds to 75% of DNA occupied with nucleosomes). DNA (146 bp) wraps around each nucleosome.

(B) A 3D DNA fiber arrangement is generated by positioning nucleosomes according to a Gaussian chain model with end-to-end distances ( $l_{e-e}$ ) calculated according to the worm like chain model (WLM) for a polymer with a persistence length of 150 bp (experimentally determined persistence length of DNA).

(legend continued on next page)



(1) nucleosomes do not form a highly ordered organization but rather arrange into discrete groups, the clutches, of various sizes and densities, which are interspaced by nucleosome-depleted regions; (2) there is a striking correlation between spatial distribution, size, and compaction of nucleosome clutches and cell pluripotency; (3) ground-state stem cells have low-density clutches containing on average only a few nucleosomes; and (4) large clutches with higher nucleosome compaction corresponds to heterochromatin and include more H1, whereas the small clutches with lower nucleosome compaction correspond to active chromatin regions since they are associated to RNA Polymerase II.

While the heterogeneity of the nucleosome clutches argues against the existence of highly ordered structures such as the 30-nm fiber, it is still possible that nucleosomes maintain an ordered organization inside the clutches. Nevertheless, our simple *in silico* model can reconstruct “nucleosome-rich” and “nucleosome-depleted” regions that recapitulate the experimental results without invoking the existence of a 30-nm fiber. Therefore, our data indicate that an ordered structure is not strictly required for the observed organization of nucleosomes.

In this work, we have discovered an important feature of embryonic stem cells, *i.e.*, their characteristic nucleosome organization along the chromatin fiber. Furthermore, we have revealed a striking correlation between naive pluripotent state and nucleosome arrangement, which was made possible by the direct visualization of nucleosomes at nanoscale resolution. We found that drugs, such as TSA, which trigger massive epigenome modifications and facilitate somatic cell reprogramming (Luis and Cosma, 2013), induce a spatial rearrangement of nucleosome clutches and modify their density. These structural modifications can potentially facilitate the maintenance of pluripotency as well as the establishment of an induced pluripotent state.

Chromatin of mESCs is hyper-dynamic, shows increased transcriptional activity and contains a high number of DNase I hypersensitivity sites (Efroni *et al.*, 2008; Fussner *et al.*, 2011b; Meshorer *et al.*, 2006; Stergachis *et al.*, 2013). These features, associated with “open” chromatin, are consistent with the existence of small, low-density clutches in mESCs. Here, by the direct visualization of nucleosomes we can now identify “open” and “closed” chromatin as small, low-density and large, high-density nucleosome clutches, respectively, and relate clutch size to cellular state. Clutch size could not only report on heterogeneities in a given mESC population but importantly, it also highly correlated with the pluripotency grade of hiPSCs.

Pluripotency grade of different hiPSCs clones can therefore potentially be characterized and compared at the single cell level using this method. Overall, these results open up exciting possibilities for identifying stem cell state simply by analyzing nucleosome arrangement. It will also be very interesting to determine whether differences in the clutches exist between different cell types such as cancer and normal cells, and if so, whether the clutch size can also be used as a diagnostic marker for cancer cell identification and consequent follow up therapies, or to identify rare subpopulations of stem/precursor cells within a specific tissue.

Nucleosome occupancy is critical for biological function since there should be a reservoir of DNA that is ready to be decoded by transcription factors and RNA polymerases. Population studies have measured an average linker-DNA length of around 50 bp between subsequent nucleosomes (Kornberg, 1977; Valouev *et al.*, 2011; Widom, 1992), which would correspond to DNA occupancy of ~75%. Here, we estimate an occupancy level of ~60% in hFbs, which might be slightly underestimated since our model does not take into account that not all nucleosomes may be labeled inside the large clutches. Our result comes very close to the occupancy level measured in genome-wide studies (Jiang and Pugh, 2009; Struhl and Segal, 2013). However, it is difficult to directly compare genome-wide chromatin immunoprecipitation (ChIP) or micrococcal nuclease (MNase) studies with STORM imaging to extract information on nucleosome number and their localization on DNA since the former methods are based on population studies and have a resolution in the range of hundreds of nanometers, whereas STORM reveals nucleosomes in single cells with much higher resolution (10–20 nm). In the future, it will be exciting to visualize both DNA and nucleosomes by STORM at specific gene loci, which may enable better comparison of the clutch data with the ChIP analysis.

## EXPERIMENTAL PROCEDURES

Full details of the experimental procedures and analyses are provided online in the [Extended Experimental Procedures](#).

### Sample Preparation and STORM Imaging

Cells were fixed with methanol-ethanol (1:1) at  $-20^{\circ}\text{C}$  for 6 min unless otherwise stated and immunostained with appropriate primary and secondary antibodies. Secondary antibodies were labeled with activator-reporter dye pairs (Alexa Fluor 405-Alexa Fluor 647) for STORM imaging. All imaging experiments were carried out with a commercial STORM microscope system from Nikon Instruments (NSTORM). Laser light at 647 nm was used for

(C) The resulting DNA fiber configuration is projected onto 2D space.

(D) In the nucleosome removal (NR) model, nucleosomes are removed from the DNA with a given probability ranging from 0 to 0.95. When a nucleosome is removed, the linker-DNA length between the neighboring nucleosomes increases by 146 bp.

(E) In the linker length (LL) model the linker-DNA lengths ( $l_i$ ) between subsequent nucleosomes are drawn from normal distributions whose averages are varied from 50 bp to 3,000 bp.

(F) Examples of synthetic STORM images obtained from the simulated arrangement of nucleosomes at 75%, 57%, and 45% nucleosome occupancy.

(G) Comparison of simulation results for the NR- (black squares and solid line) and LL-Models (white circles and dotted line) to experimental data for hFbs (horizontal blue line) and TSA-hFbs (horizontal red line) at different levels of nucleosome occupancy (x axis). The comparison is made for the number of localizations per clutch (upper), nnds of clutches (middle) and clutch area (lower). The vertical thick blue lines and black arrows show the nucleosome occupancy values for which the simulation results of the different models intersect the experimental data for the hFbs. Similarly, the vertical thick red lines and black arrows show the nucleosome occupancy values for which the simulation results intersect the experimental data for the TSA-hFbs. Trend lines are polynomial fits.

See also [Figure S6](#).

exciting Alexa Fluor 647, and laser light at 405 nm was used for activating it via an activator dye (Alexa Fluor 405)-facilitated manner. For all single color H2B imaging experiments, activation laser (405 nm) power was increased over time in an identical way according to Table S1. For dual color imaging, a second activator-reporter dye pair (Cy3-Alexa Fluor 647) and an additional activation laser at 560 nm was used. The emitted light was collected by an oil immersion 100 $\times$ , 1.49 NA objective, filtered by an emission filter (ET705/72 m), and imaged onto an electron multiplying charge coupled device (EMCCD) camera at an exposure time of 15 ms per frame. For live-cell imaging, cells were transfected with H2B-mEos2 or H2B-PAmCherry. Laser light at 405 nm was used to photoactivate the fluorescent proteins and laser light at 560 nm was used to excite the photoactivated forms. The fluorescence emission was filtered with an emission filter (BP 605/52) and recorded with an exposure time of 50 ms per frame.

### Data Analysis

STORM images were analyzed using custom-written software (Insight3, provided by Bo Huang, University of California, San Francisco) by fitting the fluorophore images in each frame to a simple Gaussian to determine x-y coordinates.

For cluster quantification, x-y localization lists were binned to construct discrete localization images with pixel size of 10 nm. These were convoluted with a square kernel ( $5 \times 5$  pixels<sup>2</sup>) to obtain density maps and transformed into binary images by applying a constant threshold. x-y coordinates in the binary image were grouped into clusters using a distance-based algorithm. Cluster sizes were calculated as the SD of x-y coordinates from the relative cluster centroid.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.054>.

### AUTHOR CONTRIBUTIONS

M.P.C., M.L., M.A.R., C.M., and M.F.G.-P. designed the experiments and data analysis. M.A.R. performed experiments and data analysis. C.M. wrote software and performed data analysis. M.P.C., M.L., M.A.R., C.M., and M.F.G.-P. wrote the manuscript. M.P.C., M.L., and M.F.G.-P. supervised the project.

### ACKNOWLEDGMENTS

We thank A. Skoultschi (Albert Einstein College of Medicine), S. Grigoryev (Penn State Hershey Cancer Institute), T. Misgeld (TUM, Munich), X. Zhuang (Harvard University) and J. Ries (European Molecular Biology Laboratory [EMBL], Heidelberg) for providing reagents and cell lines, B. Huang (University of California, San Francisco [UCSF]) for Insight3 STORM analysis software, L. Batlle Morera (Center for Genomic Regulation [CRG], Barcelona) for helping in hiPSC generation, J. Font Mateu (CRG, Barcelona) for technical suggestions on in vitro polynucleosome, F. Sottile (CRG, Barcelona) for cloning the H2B-mEos2 plasmid. We thank M. Beato, C. di Vona, G. Filion, T. Graf, B. Lehner, M. Marti-Renom, M. Mendoza, J. Sharpe, J. Solon and J. Valcarcel (CRG, Barcelona), E. Meshorer (Hebrew University of Jerusalem), A. Oddone and J. Otterstrom (Institute of Photonic Sciences [ICFO], Barcelona), and X. Zhuang (Harvard University) for critical reading of the manuscript and helpful discussions. We acknowledge the ICFO Super-resolution Light Nanoscopy (SLN) facility, the ICFO Nikon Center of Excellence in STORM, the CRG microscopy facility and the EMBL electron microscopy core facility. We thank E. Sahagún Alonso (Scixel) and A. Hirschmann (ICFO) for help with the graphical abstract. The research leading to these results was supported in part by the Fundació Cellex Barcelona (M.L.), the System's Microscopy Network of Excellence consortium (call identifier FP-7-HEALTH.2010.2.1.2.2 to M.L.), by the European Union Seventh Framework Programme under the European Research Council grants 337191-MOTOS to M.L. and 242630-RERE to M.P.C., a Human Frontier Science Program (HFSP) grant (M.P.C.), and MAT2011-22887

(M.F.G.-P). This article reflects only the author's views and the Union is not liable for any use that may be made of the information contained therein.

Received: June 4, 2014

Revised: October 10, 2014

Accepted: January 16, 2015

Published: March 12, 2015

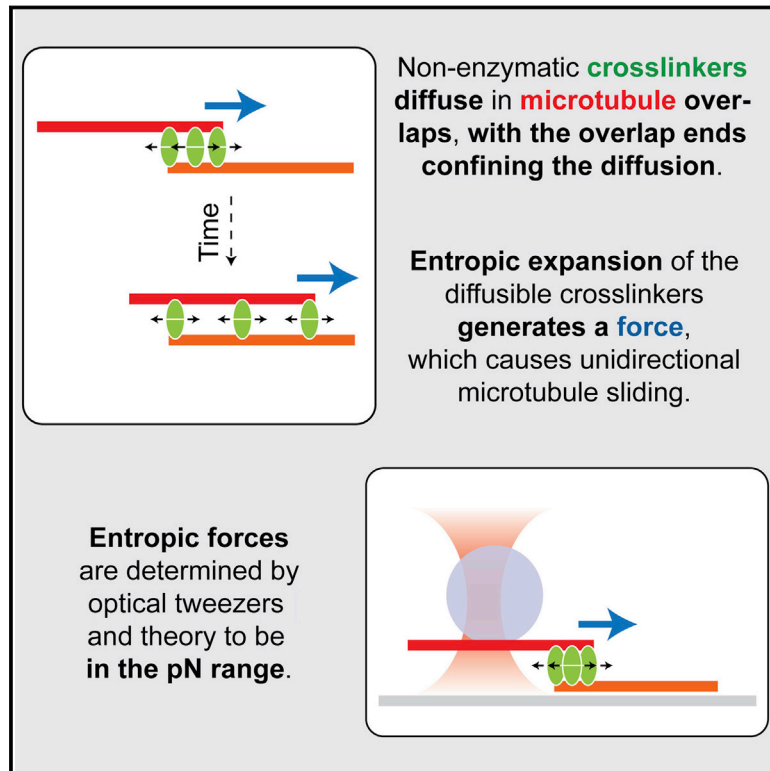
### REFERENCES

- Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 144, 439–452.
- Bohn, M., Diesinger, P., Kaufmann, R., Weiland, Y., Müller, P., Gunkel, M., von Ketteler, A., Lemmer, P., Hausmann, M., Heermann, D.W., and Cremer, C. (2010). Localization microscopy reveals expression-dependent parameters of chromatin nanostructure. *Biophys. J.* 99, 1358–1367.
- Cahan, P., and Daley, G.Q. (2013). Origins and implications of pluripotent stem cell variability and heterogeneity. *Nat. Rev. Mol. Cell Biol.* 14, 357–368.
- Clausell, J., Happel, N., Hale, T.K., Doenecke, D., and Beato, M. (2009). Histone H1 subtypes differentially modulate chromatin condensation without preventing ATP-dependent remodeling by SWI/SNF or NURF. *PLoS ONE* 4, e0007243.
- Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.* 22, 746–755.
- Dani, A., Huang, B., Bergan, J., Dulac, C., and Zhuang, X. (2010). Superresolution imaging of chemical synapses in the brain. *Neuron* 68, 843–856.
- Duricic, N., Cuervo, L.L., and Lakadamyali, M. (2014). Quantitative super-resolution microscopy: pitfalls and strategies for image analysis. *Curr. Opin. Chem. Biol.* 20, 22–28.
- Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoepfner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2, 437–447.
- Fan, Y., Nikitina, T., Zhao, J., Fleury, T.J., Bhattacharyya, R., Bouhassira, E.E., Stein, A., Woodcock, C.L., and Skoultschi, A.I. (2005). Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123, 1199–1212.
- Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. USA* 73, 1897–1901.
- Fussner, E., Ching, R.W., and Bazett-Jones, D.P. (2011a). Living without 30nm chromatin fibers. *Trends Biochem. Sci.* 36, 1–6.
- Fussner, E., Djuric, U., Strauss, M., Hotta, A., Perez-Iratxeta, C., Lanner, F., Dilworth, F.J., Ellis, J., and Bazett-Jones, D.P. (2011b). Constitutive heterochromatin reorganization during somatic cell reprogramming. *EMBO J.* 30, 1778–1789.
- Fussner, E., Strauss, M., Djuric, U., Li, R., Ahmed, K., Hart, M., Ellis, J., and Bazett-Jones, D.P. (2012). Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. *EMBO Rep.* 13, 992–996.
- Grigoryev, S.A., Arya, G., Correll, S., Woodcock, C.L., and Schlick, T. (2009). Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proc. Natl. Acad. Sci. USA* 106, 13317–13322.
- Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10, 161–172.
- Joti, Y., Hikima, T., Nishino, Y., Kamada, F., Hihara, S., Takata, H., Ishikawa, T., and Maeshima, K. (2012). Chromosomes without a 30-nm chromatin fiber. *Nucleus* 3, 404–410.
- Kamakaka, R.T., and Biggins, S. (2005). Histone variants: deviants? *Genes Dev.* 19, 295–310.

- Kornberg, R.D. (1977). Structure of chromatin. *Annu. Rev. Biochem.* 46, 931–954.
- Lluis, F., and Cosma, M.P. (2013). Resetting epigenetic signatures to induce somatic cell reprogramming. *Cell. Mol. Life Sci.* 70, 1413–1424.
- Lluis, F., Ombrato, L., Pedone, E., Pepe, S., Merrill, B.J., and Cosma, M.P. (2011). T-cell factor 3 (Tcf3) deletion increases somatic cell reprogramming by inducing epigenome modifications. *Proc. Natl. Acad. Sci. USA* 108, 11912–11917.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Marks, H., Kalkan, T., Menafra, R., Denisssov, S., Jones, K., Hofmeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590–604.
- Matsuda, A., Shao, L., Boulanger, J., Kervrann, C., Carlton, P.M., Kner, P., Agard, D., and Sedat, J.W. (2010). Condensed mitotic chromosome structure at nanometer resolution using PALM and EGFP- histones. *PLoS ONE* 5, e12768.
- Meister, P., and Taddei, A. (2013). Building silent compartments at the nuclear periphery: a recurrent theme. *Curr. Opin. Genet. Dev.* 23, 96–103.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* 10, 105–116.
- Miyazari, Y., Ziegler-Birling, C., and Torres-Padilla, M.E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. *Nat. Struct. Mol. Biol.* 20, 1321–1324.
- Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., Hihara, S., Frangakis, A.S., Imamoto, N., Ishikawa, T., and Maeshima, K. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J.* 31, 1644–1653.
- Rust, M.J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* 3, 793–795.
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 344, 376–380.
- Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell* 154, 888–903.
- Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* 20, 267–273.
- Tam, W.L., Lim, C.Y., Han, J., Zhang, J., Ang, Y.S., Ng, H.H., Yang, H., and Lim, B. (2008). T-cell factor 3 regulates embryonic stem cell pluripotency and self-renewal by the transcriptional control of multiple lineage pathways. *Stem Cells* 26, 2019–2031.
- Tóth, K.F., Knoch, T.A., Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C.P., Müller, G., and Rippe, K. (2004). Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Sci.* 117, 4277–4287.
- Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z., and Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. *Nature* 474, 516–520.
- Wendt, K.S., and Grosveld, F.G. (2014). Transcription in the context of the 3D nucleus. *Curr. Opin. Genet. Dev.* 25, 62–67.
- Widom, J. (1992). A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl. Acad. Sci. USA* 89, 1095–1099.
- Wombacher, R., Heidbreder, M., van de Linde, S., Sheetz, M.P., Heilemann, M., Cornish, V.W., and Sauer, M. (2010). Live-cell super-resolution imaging with trimethoprim conjugates. *Nat. Methods* 7, 717–719.
- Woodcock, C.L., Skoultchi, A.I., and Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res.* 14, 17–25.
- Wray, J., Kalkan, T., and Smith, A.G. (2010). The ground state of pluripotency. *Biochem. Soc. Trans.* 38, 1027–1032.
- Yi, F., Pereira, L., and Merrill, B.J. (2008). Tcf3 functions as a steady-state limiter of transcriptional programs of mouse embryonic stem cell self-renewal. *Stem Cells* 26, 1951–1960.
- Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–523.

# Diffusible Crosslinkers Generate Directed Forces in Microtubule Networks

## Graphical Abstract



## Authors

Zdenek Lansky, Marcus Braun, ..., Marcel E. Janson, Stefan Diez

## Correspondence

tenwolde@amolf.nl (P.R.t.W.), marcel.janson@wur.nl (M.E.J.), diez@bcube-dresden.de (S.D.)

## In Brief

Mechanical force on the same scale as force induced by microtubule-crosslinking motor proteins can be generated in microtubule networks by passive diffusion of crosslinkers in confined spaces between cytoskeletal filaments

## Highlights

- Diffusible crosslinkers confined in microtubule overlaps can generate directed forces
- Thermal motion of the crosslinkers is harnessed to perform mechanical work
- Force generation is quantitatively explained by entropic expansion of crosslinkers





# Diffusible Crosslinkers Generate Directed Forces in Microtubule Networks

Zdenek Lansky,<sup>1,2,5</sup> Marcus Braun,<sup>1,2,5</sup> Annemarie Lüdecke,<sup>1</sup> Michael Schlierf,<sup>1</sup> Pieter Rein ten Wolde,<sup>3,\*</sup> Marcel E. Janson,<sup>4,\*</sup> and Stefan Diez<sup>1,2,\*</sup>

<sup>1</sup>B CUBE - Center for Molecular Bioengineering, Technische Universität Dresden, Arnoldstrasse 18, 01307 Dresden, Germany

<sup>2</sup>Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany

<sup>3</sup>AMOLF, Science Park 104, 1098 XG Amsterdam, the Netherlands

<sup>4</sup>Laboratory of Cell Biology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

<sup>5</sup>Co-first author

\*Correspondence: [tenwolde@amolf.nl](mailto:tenwolde@amolf.nl) (P.R.t.W.), [marcel.janson@wur.nl](mailto:marcel.janson@wur.nl) (M.E.J.), [diez@bcube-dresden.de](mailto:diez@bcube-dresden.de) (S.D.)

<http://dx.doi.org/10.1016/j.cell.2015.01.051>

## SUMMARY

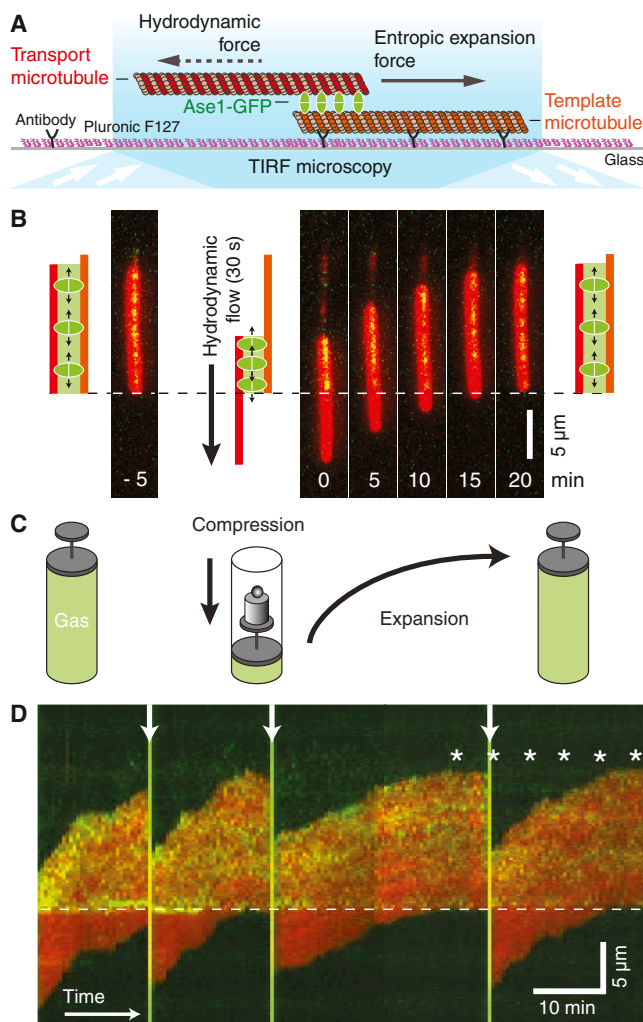
Cytoskeletal remodeling is essential to eukaryotic cell division and morphogenesis. The mechanical forces driving the restructuring are attributed to the action of molecular motors and the dynamics of cytoskeletal filaments, which both consume chemical energy. By contrast, non-enzymatic filament crosslinkers are regarded as mere friction-generating entities. Here, we experimentally demonstrate that diffusible microtubule crosslinkers of the Ase1/PRC1/Map65 family generate directed microtubule sliding when confined between partially overlapping microtubules. The Ase1-generated forces, directly measured by optical tweezers to be in the piconewton-range, were sufficient to antagonize motor-protein driven microtubule sliding. Force generation is quantitatively explained by the entropic expansion of confined Ase1 molecules diffusing within the microtubule overlaps. The thermal motion of crosslinkers is thus harnessed to generate mechanical work analogous to compressed gas propelling a piston in a cylinder. As confinement of diffusible proteins is ubiquitous in cells, the associated entropic forces are likely of importance for cellular mechanics beyond cytoskeletal networks.

## INTRODUCTION

Diffusion, originating from the random, thermal motion of molecules, is one of nature's most important transport mechanisms. It can be exploited for the generation of directed forces when the molecules are spatially confined. An every-day example is a gas spring, where the expansion of a gas compressed in a cylinder can be understood as an entropy-driven process that maximizes the total number of microscopic states the system can adopt. We here ask if, analogously, subcellular mechano-systems like cytoskeletal networks, can harness the entropic forces arising from the confinement of diffusible molecules.

In many cellular systems, the molecules are not confined to three dimensions, but rather to two dimensions or even to one dimension. A prominent example of the latter is the diffusion of proteins along microtubules (Helenius et al., 2006). Moreover, the ends of microtubules have been shown to constitute diffusion barriers for proteins involved in forcefully tethering kinetochores to the shrinking ends of depolymerizing microtubules (Asbury et al., 2006; Gestaut et al., 2008; Powers et al., 2009), as well as for diffusible microtubule crosslinkers (Braun et al., 2011). An example of a diffusible microtubule crosslinker is *S. pombe* Ase1 (a member of the Ase1/PRC1/Map65 family), which is believed to stabilize bipolar microtubule arrays. Ase1 localizes to the anti-parallel microtubule overlaps in the midzone of the mitotic spindle during anaphase (Yamashita et al., 2005) and to the anti-parallel microtubule overlaps of the interphase microtubule array (Loiodice et al., 2005). While forces generated by molecular motors and dynamic microtubules are believed to be the main contributors to the remodeling of both of these bipolar microtubule structures (Civelekoglu-Scholey and Scholey, 2010; Janson et al., 2007; Peterman and Scholey, 2009), bipolar microtubule arrays are destabilized and break down in the absence of Ase1 (Loiodice et al., 2005; Schuyler et al., 2003; Yamashita et al., 2005). Since Ase1 crosslinkers slow down microtubule-microtubule sliding (Braun et al., 2011; Janson et al., 2007), friction forces by microtubule-bound Ase1 may thus be required to balance motor forces within networks. Still, because Ase1 can diffuse in the confined space of microtubule overlaps (Braun et al., 2011; Kapitein et al., 2008), we reasoned that Ase1, apart from generating friction, might also generate entropic forces.

We here devised a well-controlled experimental assay to confine small numbers of Ase1 diffusible crosslinkers in between two partially overlapping microtubules. Using total-internal reflection fluorescence (TIRF) microscopy, we showed that the entropic expansion of the confined crosslinkers is strong enough to induce the directed sliding of the microtubules with respect to each other. We directly measured the entropic forces generated by Ase1 in the expanding overlaps using optical tweezers and found them to be in the piconewton (pN) range. This suggests that the entropy of the crosslinkers in an overlap can generate biologically relevant forces that are on the same scale as forces induced by microtubule-crosslinking motor proteins. To test this hypothesis, we employed kinesin-14 motor proteins and found



**Figure 1. Entropic Expansion of Diffusible Ase1-GFP Crosslinkers Induces the Directed Sliding of Partially Overlapping Microtubules**

(A) Schematic representation of Ase1-driven sliding of a transport microtubule (red) along a surface-immobilized template microtubule (orange).

(B) Typical time-lapse fluorescence, multichannel micrographs showing the positions of a transport microtubule (red channel) as a function of time before and after flow-induced compression of Ase1-GFP (green channel) within a microtubule overlap. Prior to imaging, free Ase1-GFP was removed from solution. Schematic diagrams illustrate the positions of the microtubules before and immediately after the application of the hydrodynamic flow, as well as at the end of the experiment. The end of the template microtubule is indicated by the dashed line.

(C) A gas spring, the macroscopic analog of the molecular Ase1-microtubule system, expands when the external load is decreased.

(D) Extended kymograph showing multiple cycles of the experiment described in (B). Time points and direction of flow application are indicated by the vertical arrows. Asterisks indicate the time of the snapshots presented in (B). The end of the template microtubule is indicated by the dashed line. Regions with enhanced localization of Ase1-GFP signal correspond to the microtubule overlap. See also [Movie S1](#).

that motor-driven microtubule-microtubule sliding could indeed be reversed by the addition of Ase1. We quantitatively describe the force generation by Ase1 by a statistical-mechanical model,

which predicts the expansion force to follow the ideal gas law. Taken together, our results show that Ase1 diffusible crosslinkers confined between partially overlapping microtubules create a pressure, analogously to gas molecules confined in a cylinder by a piston. Our results are a demonstration of the unexpected effects entropy may have in cells. We suggest that forces generated by diffusible crosslinkers of the Ase1/PRC1/MAP65 family are likely of importance in the midzone of the mitotic spindle, where they may regulate the motorized sliding of anti-parallel microtubules.

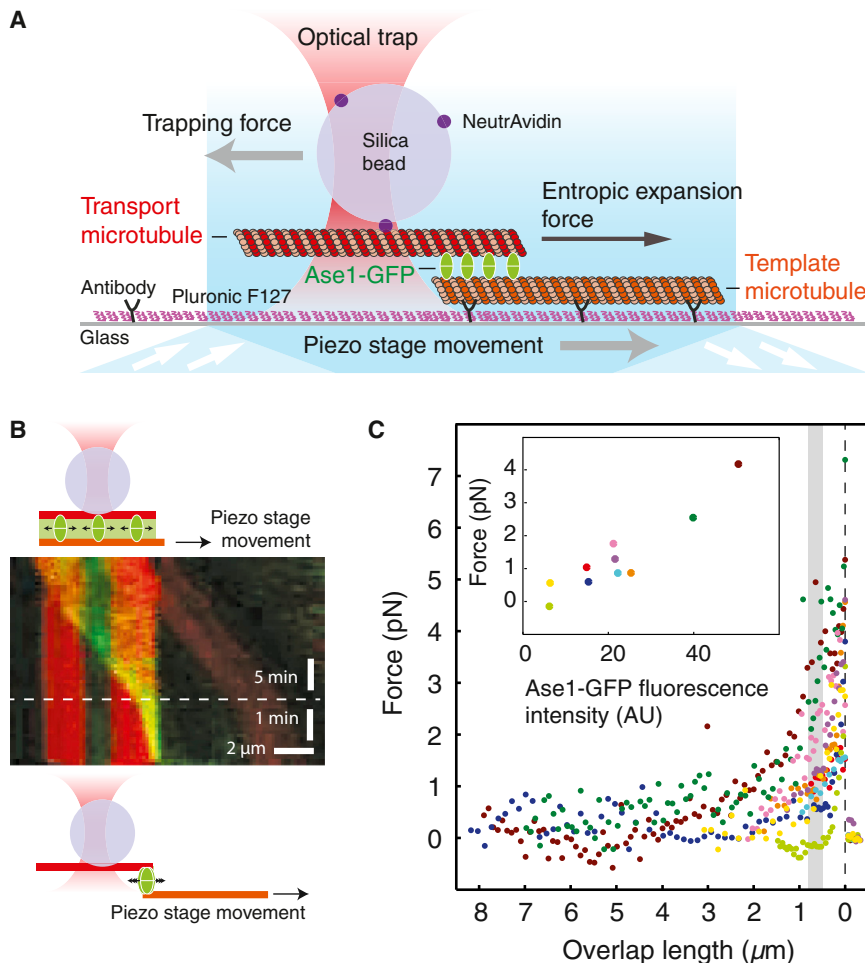
## RESULTS

### Entropic Expansion of Diffusible Ase1-GFP Crosslinkers Induces the Directed Sliding of Partially Overlapping Microtubules

To study force generation by confined Ase1 crosslinkers in vitro, we generated overlapping microtubules by (1) immobilizing dimly rhodamine-labeled “template” microtubules on a coverslip, (2) allowing 50 picomolar (pM) Ase1-GFP to bind diffusively to the immobilized template microtubules, and (3) flushing in brightly rhodamine-labeled “transport” microtubules to bind to the template microtubules using a solution without Ase1-GFP; this effectively removed Ase1-GFP molecules that were not bound to the template microtubules ([Figure 1](#); [Experimental Procedures](#)). We then applied hydrodynamic flow of assay buffer without Ase1-GFP to slide the transport microtubules along the template microtubules, generating partial overlaps with reduced overlap lengths ([Figures 1A](#) and [1B](#)). Due to their high affinity for microtubule overlaps, as compared to their lower affinity for single microtubules ([Braun et al., 2011](#)), the diffusible Ase1-GFP molecules did not leave the overlap regions during this process. The reduction in the overlap lengths consequently led to an increased confinement of the crosslinkers. As soon as the flow stopped, the overlap lengths increased through directed sliding of the transport microtubules ([Figure 1B](#) and [Movie S1](#)). During this expansion, the confined Ase1-GFP molecules redistributed themselves uniformly within the overlap regions by one-dimensional diffusion. Again, no Ase1-GFP molecules were lost as evidenced by the constancy of the integrated Ase1-GFP fluorescence intensity along the overlap regions ([Figure S1B](#)). Compression and expansion could be cyclically repeated ([Figure 1D](#) and [Movie S1](#)), resembling the macroscopic mechanism of a gas spring ([Figure 1C](#)).

### Quantification of the Forces Generated by Ase1-GFP Confined between Partially Overlapping Microtubules Using Optical Tweezers

We quantified the forces generated by Ase1 confined between partially overlapping microtubules by optical tweezers ([Figure 2](#)). First, we formed microtubule overlaps in a similar manner as in the previous experiment ([Experimental Procedures](#)). In the absence of Ase1-GFP in solution, we attached a silica microsphere to a transport microtubule by optical tweezers. Using a piezo translation stage, we then moved the template microtubule in steps relative to the laser trap in the direction along the longitudinal axis of the template microtubule, forming partial microtubule-overlaps and compacting Ase1-GFP until the two microtubules were pulled apart ([Figures 2A](#), [2B](#), [S2A](#), and [S2B](#)



**Figure 2. Quantification of the Entropic Forces Generated by Ase1-GFP Confined between Partially Overlapping Microtubules by Optical Tweezers**

(A) Schematic representation of the optical tweezers experiment. A trapped, NeutrAvidin-coated silica bead (not drawn to scale) is attached to a biotinylated transport microtubule (red). In order to slide the microtubules relative to each other, the template microtubule (orange) was moved by a piezo translational stage, while keeping the center of the laser trap at a fixed position.

(B) Typical multichannel kymograph showing the movement of the dimly labeled template microtubule (driven by the movement of the piezo stage) relative to the trapped, brightly labeled transport microtubule in the absence of free Ase1-GFP in solution. The density of Ase1-GFP increased in the shortening overlap. Approximately 2 min before the separation of the microtubules, the movement of the piezo stage was slowed down to obtain a higher number of data points. The bleached spot in the middle of the transport microtubule is caused by the focused trapping laser. The region with enhanced localization of Ase1-GFP signal corresponds to the microtubule overlap. For snapshots of the event see [Figure S2A](#). See also [Movie S2](#).

(C) Equilibrium bead displacements, corresponding to the steady-state forces induced by the confined Ase1-GFP in the overlaps, as function of overlap length. Presented are ten independent measurements. The inset shows the measured forces as function of Ase1-GFP fluorescence intensity in the overlap averaged for overlaps with lengths between 0.6 and 0.8  $\mu\text{m}$  (denoted by the gray box in the main panel; same color-coding of measurements). Overlap lengths and forces were offset-corrected by assuming that the overlap length is zero right before the microtubules were pulled apart (dashed line) and that the force is zero after the microtubules were pulled apart.

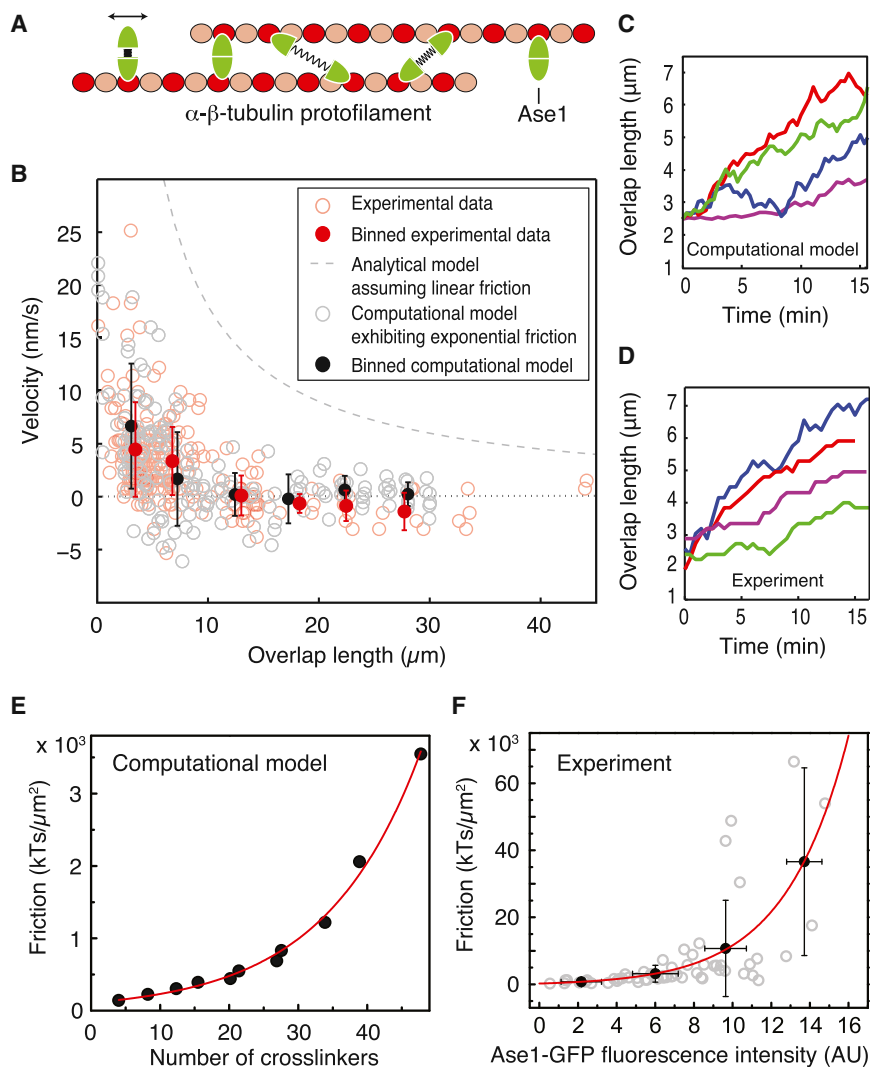
and [Movie S2](#)). After each step, we allowed the system to equilibrate before measuring the force. We found that the force increased with decreasing overlap length, reaching values up to  $3.7 \pm 1.8$  pN (average  $\pm$  SD,  $n = 10$ ) just before the two microtubules were pulled apart ([Figure 2C](#)). The observed forces increased linearly with increasing Ase1-GFP densities in the overlaps as inferred from fluorescence intensities ([Figures 2C](#), inset, and [S2C](#); [Experimental Procedures](#)).

### Modeling the Ase1-Induced Expansion of Partial Microtubule-Overlaps

To explain the origin of the observed forces generated by Ase1, we analytically modeled the mutually exclusive binding of crosslinkers to discrete binding sites along a single protofilament in a microtubule overlap ([Figure 3](#)). For the case of a constant number of confined crosslinkers in the overlap; i.e., when no crosslinkers bind into or unbind from the overlap (scenario as in [Figures 1](#) and [2](#)), the entropic expansion force,  $F$ , acting on the transport microtubule is found to be given by the one-dimensional analog of the ideal gas law  $FL \cong nk_B T$  ([Extended Results](#), Text 1). Here,  $L$  is the overlap length and  $n$  is the number of

crosslinkers within the overlap,  $k_B$  is the Boltzmann constant, and  $T$  is the absolute temperature. This model predicts that the force increases linearly with the density of the crosslinkers in the overlap, as observed in our experiments ([Figures 2C](#), inset, and [S2C](#)). While a quantitative test of the predicted relation between force and crosslinker density is not possible due to experimental uncertainties in overlap lengths and protein numbers, the range of maximum measured forces is predicted correctly. The model predicts the generation of forces in the 1 pN range when the crosslinkers are maximally compressed between two microtubule protofilaments, that is when all binding sites within the overlap are fully occupied by Ase1. Structural work on Ase1 homologs suggests that such high densities of crosslinkers are indeed possible ([Subramanian et al., 2010](#)). The observed maximal forces of  $3.7 \pm 1.8$  pN may indicate that multiple rows of Ase1 crosslinkers bind to neighboring protofilaments in the overlap ([Extended Results](#), Text 1).

We next investigated whether entropic forces, in combination with frictional drag exerted by the Ase1 crosslinkers, can also explain the observed sliding velocities of transport microtubules



**Figure 3. Ase1-GFP Entropy Together with an Exponential Scaling of Friction Explain the Expansion of Microtubule Overlaps in the Absence of Ase1-GFP in Solution**

(A) Schematic representation of the modeled geometry. Microtubules are modeled as a one-dimensional array of lattice sites. For the computational model, microtubule-Ase1-microtubule links are simulated as harmonic springs, whose ends can hop individually between neighboring lattice sites. The spring constant is chosen to match the measured diffusion rates of Ase1-GFP on single microtubules and in microtubule overlaps. Rate constants for crosslinker binding and unbinding are in agreement with the measured dissociation constants (Figures S3A and S3B; Extended Results, Texts 1–4; Table S1).

(B) Averaged velocities of Ase1-GFP induced microtubule sliding as function of overlap length. Shown are experimental data (red open circles, 95 events, 48 microtubules in experiments as presented in Figure 1B), results from the analytical model (gray dashed line,  $v_{MT} = 2 D_{Ase1}^{MT} / L$  with  $D_{Ase1}^{MT} = 0.085 \pm 0.007 \mu m^2 s^{-1}$  assuming a constant number of Ase1-GFP in the overlap), as well as results from the computational model (gray open circles, total of 24 simulation runs, parameters summarized in Table S1). In the computational model, the initial number of crosslinkers  $n_0$  and initial overlap lengths  $L_0$  were chosen from the experimentally observed range of  $n_0 = 10, 20, 50$ , and  $L_0$  randomly between 0.1 and 30  $\mu m$ , respectively. The overlaps were allowed to expand for at least 15 min. Solid red and black circles represent the binned averages ( $\pm SD$ ) of the experimental data and the computational model, respectively. Data points (overlap lengths ranging from 0 to 30  $\mu m$ ) were binned in six equidistant bins with a width of 5  $\mu m$ .

(C and D) Typical time traces of overlap expansions obtained from the computational model (C, data as shown in Movie S3 and summarized in Figure 3B) and from the experiments (D, data as shown in Movie S4 and summarized in Figure 3B).

shown in Figure 1B and Movie S1 and summarized in Figure 3B). Different colors represent individual events. The variability in the time traces reflects the stochasticity of the underlying force-generating mechanism.

(E) Results of the computational model predicting that friction increases exponentially with the number of crosslinkers. The diffusion constants of transport microtubules on an infinitely long template microtubule were determined by computing their mean-square displacements as a function of time, for different numbers of crosslinkers. Friction coefficients  $\gamma$  were calculated from the computed diffusion coefficients  $D$  using  $\gamma = k_B T / D$ . The simulation parameters are summarized in Table S1.

(F) Experimental results showing that friction increases exponentially with the number of Ase1-GFP crosslinkers, as inferred from the Ase1-GFP fluorescence intensity integrated along the overlap region. Friction coefficients were calculated from the diffusion of single transport microtubules (Figure S3D and Movie S4) by the same procedure as in (C). Friction coefficients ( $n = 56$  diffusing microtubules) were binned according to the Ase1-GFP fluorescence intensity measured in the overlap during the movement into four equidistant bins with the width of 4 AU (solid black circles represent averages  $\pm SD$ ).

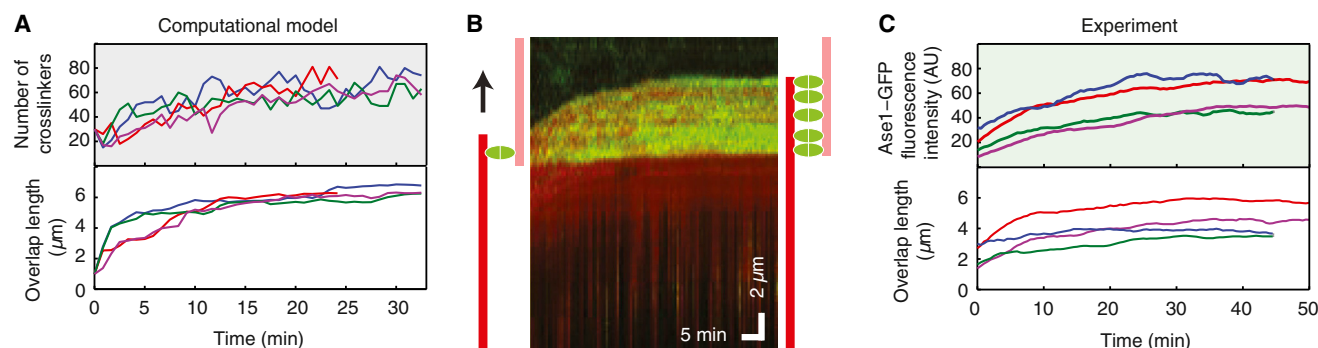
in absence of external load (scenario as in Figure 1). The viscous drag exerted by the solution was neglected due to its small contribution at low velocities (Hunt et al., 1994; Tawada and Sekimoto, 1991). We described the frictional drag coefficient  $\gamma$  of a single Ase1-microtubule link following the Einstein relation  $\gamma = k_B T / D_{Ase1}^{MT}$  (Einstein, 1906), where  $D_{Ase1}^{MT}$  is the diffusion constant of a single Ase1 molecule on a single microtubule. Assuming a linear dependence of the frictional drag on the number of diffusible crosslinkers (Tawada and Sekimoto, 1991), the velocity of overlap expansion is given by  $v_{MT} = 2 D_{Ase1}^{MT} / L$

(Extended Results, Text 2). This analytical expression, independent of the number of crosslinkers in the overlap, qualitatively reproduced the trend of the measured velocities (Figure 3B). However, it overestimated the absolute values, suggesting that friction might be underestimated in our analytical model.

### Friction between Crosslinked Microtubules Depends Exponentially on the Number of Crosslinkers

To explain the magnitude of entropy-driven sliding velocities quantitatively, we set up a particle-based, computational model





**Figure 4. Overlap Expansion Slows Down in the Presence of Ase1-GFP in Solution**

(A) Typical simulated time traces (computational model) of overlap length and number of crosslinkers in the overlap during microtubule sliding in the presence of Ase1 in solution. At 0.1 nM Ase1 concentration, overlap expansion comes to an apparent stall before reaching full overlap (20 μm in this particular case). Model parameters are listed in Table S1; the initial number of crosslinkers was 30. Different colors represent individual simulated events. The variability in the time traces reflects the stochasticity of the underlying mechanism.

(B) Typical multichannel kymograph of Ase1-GFP driven sliding of a transport microtubule (red) in the presence of 17 pM Ase1-GFP (green) in solution. In contrast to experiments with a constant number of Ase1-GFP in the overlap (see Figures 1 and 3), sliding comes to a halt due to the increase in the number of Ase1-GFP molecules, and thus the Ase1-GFP induced friction, in the expanding overlap. The region with enhanced localization of Ase1-GFP signal corresponds to the microtubule overlap.

(C) Typical experimental time traces (out of 15 captured events) of overlap length and number of crosslinkers (data as presented in the kymograph in Figure 4B). Different colors represent individual events. The variability in the time traces reflects the stochasticity of the underlying mechanism.

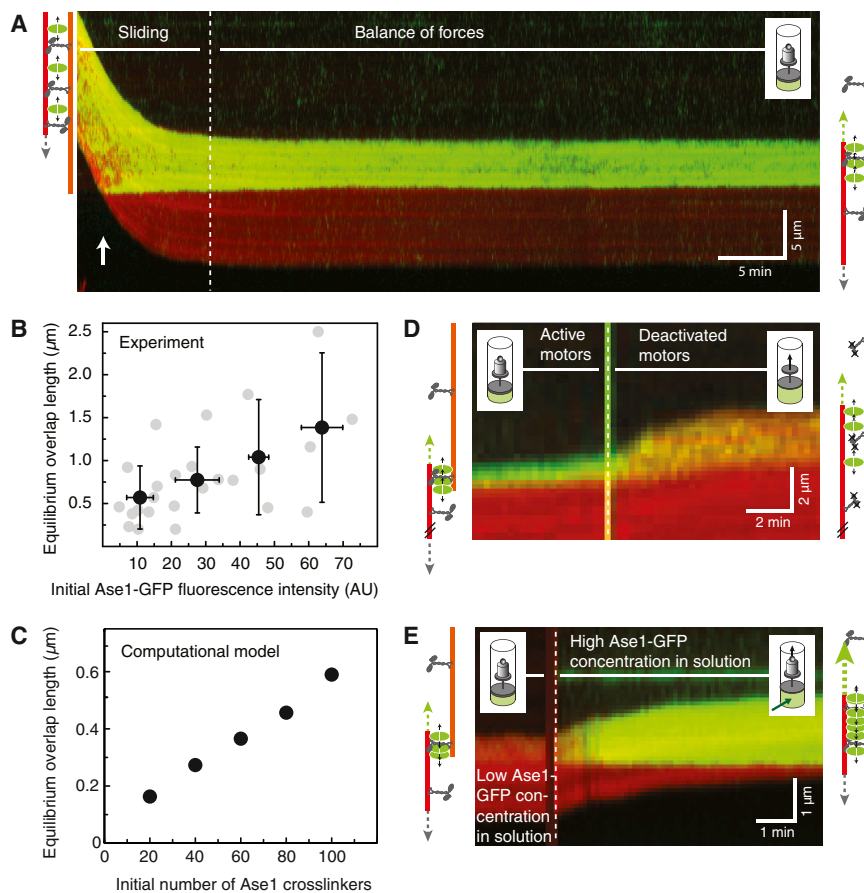
in which the microtubule-Ase1-microtubule links are described as harmonic springs whose ends can hop between neighboring binding sites within an overlap formed by two opposing protofilaments (Figure 3A; Extended Results, Text 3). The spring constant of the individual microtubule-Ase1-microtubule links was estimated from the about 8-fold lower diffusion coefficient of Ase1-GFP in microtubule overlaps compared to Ase1-GFP on single microtubules (Figure S3C). This computational model yielded time traces of the overlap expansion that are in good agreement with those measured experimentally (Figures 3B–3D and Movies S1 and S3). Interestingly, our computational model predicted that the total friction between two microtubules increases exponentially, instead of linearly, with the number of Ase1 crosslinkers in the overlap (Figure 3E). This non-linearity explains why our simple analytical model overestimated the sliding velocities.

To test the predicted exponential dependence of the friction on the number of crosslinkers, we characterized the Ase1-generated friction experimentally. We again formed microtubule overlaps in a similar manner as in the previous experiments (Experimental Procedures). In the absence of Ase1-GFP in solution, we observed the transport microtubules diffusing along the template microtubules (Figure S3D and Movie S4). For each transport microtubule that fully overlapped with a template microtubule, we estimated the diffusion coefficient by determining the mean square displacement as a function of time. Using the Einstein relation, we calculated the friction between the transport and the template microtubules. While a quantitative comparison between simulations and experiments was not possible due to experimental uncertainties in determining the number of crosslinkers (Experimental Procedures), the experiments confirmed the predicted exponential increase of the friction with the number of Ase1-GFP molecules as inferred from fluorescence intensities (Figure 3F).

### Ase1-GFP Condensation Slows Down the Overlap Expansion

Experiments so far were performed in the absence of Ase1 in solution. In this situation, the entropic force for overlap expansion decreased with overlap length, as the available Ase1 was diluted in the overlap (Extended Results, Text 1). However, in the presence of Ase1 in solution, the binding of new crosslinkers into filament overlaps, “crosslinker condensation”, generates additional forces for filament sliding (Lan et al., 2009; Peskin et al., 1993; Zandi et al., 2003). Our analytical model predicts that, in the presence of crosslinker condensation, a constant, length-independent, driving force is obtained, analogous to gas being continuously added into an expanding gas spring such that the pressure remains constant (Extended Results, Text 1). However, binding of additional crosslinkers also increases friction. The analytical model, in which the friction increases linearly with the number of bound crosslinkers, predicts that condensation of new crosslinkers to the expanding overlap slightly increases the velocity of overlap expansion, as compared to the case in which overlap expansion is driven by entropy alone (Extended Results, Text 1). In contrast, the computational model, in which the friction increases exponentially with the number of bound crosslinkers, predicts that binding of new crosslinkers to the overlap rapidly brings overlap expansion to a standstill; while crosslinker condensation keeps the driving force for overlap expansion constant, the exponential increase of the friction prohibits sliding (Figure 4A).

To test this prediction experimentally, we added assay buffer with Ase1-GFP to partially overlapping microtubules. This resulted in Ase1-GFP condensation into the overlap. We observed that the microtubules slid slower as compared to the situation without crosslinker condensation and that, in agreement with our computational model, they stopped sliding before full overlap was reached (Figures 4B, 4C, and S4). Our computational model



**Figure 5. Ase1-Induced Entropic Forces Balance the Forces Exerted by Multiple Ncd Motors**

(A) Typical multichannel kymograph showing the sliding of a transport microtubule (red) on top of an immobilized template microtubule. Partial microtubule-overlaps were formed in the presence of 312 pM Ase1-GFP (green) and 300 pM Ncd, resulting in a force equilibrium between the Ncd-motor generated force (acting in the direction of decreasing overlap length) and the Ase1-GFP entropic force (acting in the direction of increasing overlap length). The region with enhanced localization of Ase1-GFP signal corresponds to the microtubule overlap. See also [Movie S5](#), left.

(B) At a constant Ncd concentration, the equilibrium overlap length increased with increasing amounts of Ase1-GFP in the overlap (Pearson's correlation coefficient = 0.6,  $p = 0.004$ ). The length of microtubule overlaps was measured in events as presented in [Figures 5A](#) and [S5B](#), at the moment when sliding had stopped ( $n = 25$  events). Ase1-GFP fluorescence intensity was integrated along the overlap region at the moment when microtubules started to separate (denoted by vertical white arrow in the kymograph in [A](#)). Solid black circles represent the binned averages ( $\pm$ SD) of the experimental data. Data points were binned in four equidistant bins with a width of 20 AU.

(C) Results of the computational model showing the positive correlation between the initial number of crosslinkers (when the microtubules start to separate) and the equilibrium overlap length. The lengths of microtubule overlaps were determined in events as presented in [Figure S5C](#), at the moment when sliding had effectively stopped. The

motor force was modeled as an external load on the transport microtubule that scales linearly with the overlap length. The simulation parameters are summarized in [Table S1](#).

(D) Typical multichannel kymograph (out of a total of six recorded events) demonstrating the shift of the force balance after deactivation of the Ncd-motors by exchanging ATP with ADP in the assay buffer. Driven by the entropic expansion of the Ase1-GFP molecules bound to the overlaps, the transport microtubule did slide in the direction of increasing overlap length. During the expansion phase, the Ncd-motor concentration was kept constant in solution (300 pM) and no free Ase1-GFP was present in solution. The region with enhanced localization of Ase1-GFP signal corresponds to the microtubule overlap. See also [Movie S5](#), middle.

(E) Typical multichannel kymograph (out of a total of eight recorded events) demonstrating the shift of the force balance after increasing the Ase1-GFP concentration in solution (from 91 pM to 1,400 pM) leading to an increase in the number of Ase1-GFP molecules binding into the overlap. The transport microtubule slid in the direction of increasing overlap length; i.e., against the ATP-driven force of Ncd (kept at constant concentration of 300 pM in solution). The region with enhanced localization of Ase1-GFP signal corresponds to the microtubule overlap. See also [Movie S5](#), right.

thus explains overlap expansion in absence ([Figures 3B–3D](#)) and presence of crosslinker condensation ([Figure 4](#)).

### Ase1-Induced Entropic Forces Balance the Forces Exerted by Multiple Microtubule-Crosslinking Motors

To test whether the forces associated with the entropic expansion of the Ase1-GFP molecules are sufficient to counteract forces generated by microtubule-crosslinking motor proteins, we formed and imaged microtubule overlaps in the presence of Ase1-GFP and *D. melanogaster* kinesin-14 Ncd ([Figure 5](#); [Experimental Procedures](#)). Ncd, which does not directly interact with Ase1 ([Braun et al., 2011](#); [Figure S5A](#)) started to slide the microtubules apart, thereby compressing the Ase1 molecules in the shortening microtubule overlaps. During this compression, the number of bound Ase1-GFP linkers stayed roughly constant because of their high affinity for the overlap, while the number of Ncd molecules decreased linearly with decreasing overlap

length ([Braun et al., 2011](#)). After about 10 min, sliding came to a halt and the lengths of the overlaps stayed constant ([Figures 5A](#) and [S5B](#) and [Movie S5](#), left). This suggests that the sliding force induced by the Ncd motors is balanced by the entropic expansion force of Ase1, analogous to a gas spring, in which the external load is balanced by the internal pressure of the gas. The equilibrium overlap lengths increased with increasing numbers of Ase1-GFP molecules in the overlap ([Figure 5B](#)). When simulating the motor force as an external load that scales linearly with overlap length ([Braun et al., 2011](#); [Furuta et al., 2013](#)), our computational model qualitatively reproduced both the establishment of an equilibrium state in which the overlap length becomes constant ([Figure S5C](#)) and the correlation between the number of crosslinkers in the overlap and the length of the overlap ([Figure 5C](#); Model parameters summarized in [Table S1](#)). A quantitative comparison between experiment and simulation was not possible due to the experimental uncertainty in the number of crosslinkers

(Experimental Procedures) and the lack of information about the magnitude and scaling of the forces generated by multiple motors (Furuta et al., 2013; Nelson et al., 2014).

In line with the hypothesis that the Ncd sliding forces are balanced by the Ase1 generated forces, we found that the overlap lengths immediately increased when either (1) Ncd motors were deactivated by exchanging ATP for ADP in the assay buffer (Figure 5D and Movie S5, middle) or (2) crosslinkers were added into the overlaps by increasing the Ase1-GFP concentration in solution (Figure 5E and Movie S5, right). Hence, just like a gas spring, the overlap expanded when the force balance was tipped by either (1) reducing the opposing, external load or (2) raising the internal pressure by increasing the number of molecules in the overlap. These findings demonstrate that diffusible crosslinkers are capable of generating entropic expansion forces of the same order of magnitude as the forces generated by multiple molecular motors.

## DISCUSSION

### Diffusible Microtubule Crosslinkers Can Generate Entropic Forces in the pN Range

Previously, cytoskeletal re-organization has been attributed to forces either generated by molecular motors (e.g., motor-driven filament sliding in muscles, mitotic spindles, or flagella) or filament dynamics (e.g., polymerization-dependent protrusions in cell motility or depolymerization-dependent chromosome segregation). In our work, we described an additional force-generating mechanism, which is based on the entropy of diffusible crosslinkers confined between partially overlapping cytoskeletal filaments. So far, the diffusion of proteins along cytoskeletal filaments has been mostly associated with the generation of mechanical friction in response to sliding movement (Bormuth et al., 2009; Braun et al., 2011; Forth et al., 2014; Janson et al., 2007; Subramanian et al., 2010). However, diffusion inside a confined space also creates a pressure that can manifest itself as a directed entropic force even in the absence of other forces. Notably, this mechanism is different from mechanisms based on condensation, which have previously been put forward as alternative force-generating mechanisms in a number of different contexts including thermal ratchets (Gayathri et al., 2012; Hill, 1985; Lan et al., 2009; Neujahr et al., 1997; Peskin et al., 1993; Sun et al., 2010; Zandi et al., 2003). Employing the diffusible microtubule crosslinker Ase1, we demonstrated the entropic force generation against three different external forces, i.e., originating from hydrodynamic flow, optical tweezers, and molecular motors. In line with the prediction of our analytical model, optical tweezers measurements revealed that the entropic forces were in the pN range when the binding sites within the overlaps were highly occupied by the Ase1-GFP molecules (Figure 2C; Extended Results). In agreement with an entropic driving force, we observed a linear increase of force with crosslinker density (Figures 2C, inset, and S2C).

### Entropic Forces Are High Enough to Balance the Forces of Motor Proteins

The crosslinker-induced forces observed in our experiments and described by our models are comparable to the forces gener-

ated by multiple molecular motors (Figure 5). Molecular motors that regulate the length of the midzone of the mitotic spindle slide overlapping microtubules apart, thereby decreasing the lengths of the microtubule overlaps (Fink et al., 2009; Kapitein et al., 2005). Our work suggests that this motion will compact the crosslinkers that are localized in the midzone overlaps (Schuyler et al., 2003; Yamashita et al., 2005), generating entropic forces that oppose the motor-driven sliding. For kinesin-14 Ncd, which is capable of exerting additive forces of about 0.1 pN per motor (Furuta et al., 2013), we were able to directly show the balance between motor forces and entropic forces. For stronger motors involved in spindle organization, such as kinesin-5, which generates forces of 5–7 pN per motor (Valentine et al., 2006), the entropic forces may not be able to fully antagonize the motor forces. However, forces generated by multiple molecular motors do not necessarily add up (Furuta et al., 2013). Moreover, recent work on Cin8 (yeast kinesin-5) shows that motors can switch directionality (Roostalu et al., 2011). Thus, the force generated by multiple kinesin-5 motors might be much lower than the simple sum of the maximal forces that are generated by each individual motor. In such cases, and in situations where motors of different types compete with one another (Hentrich and Surrey, 2010), Ase1 could play a major role in setting the force balance. Ase1-induced entropic forces may indeed help to stabilize overlaps in the spindle midzones during mitosis, where ensembles of molecular motors are involved in the control of microtubule sliding.

Our results suggest that entropic expansion and condensation of Ase1, as well as of other diffusible microtubule crosslinkers, constitute an additional layer of regulation for the dynamic control of microtubule overlap length, besides the regulation of microtubule dynamics (Bieling et al., 2010) and force production by opposing molecular motors (Hentrich and Surrey, 2010). Phospho-regulation of Ase1 during the cell cycle (Fu et al., 2009) could regulate the difference in affinity of crosslinkers for microtubule overlaps and for single microtubules. This would enable control over the magnitudes of the entropic and condensation forces.

In the future, it will be interesting to study entropic force generation with proteins from other organisms. Although the vertebrate Ase1 homolog PRC1 unbinds faster from microtubules than Ase1, similarly to Ase1, PRC1 has a preference for binding to microtubule overlaps compared to binding to single microtubules (Bieling et al., 2010). PRC1 is thus also confined between overlapping microtubules, and since it is diffusible, is likely to generate entropic forces. Extensive *in vivo* work will be necessary to determine the magnitude of the crosslinker-induced forces in the different scenarios. Besides the generation of forces between anti-parallel microtubules, diffusible crosslinkers may also exert forces between parallel microtubules. In such a geometry, where crosslinking motors fail to generate directed motion (Braun et al., 2009; Fink et al., 2009; Kapitein et al., 2005), overlap maximization may aid the focusing of parallel microtubules into poles in the absence of centrosomes (Compton, 1998).

### Crosslinker Friction Scales Exponentially with Crosslinking Number

When new crosslinkers condense into the expanding overlap, the friction between the microtubules rises. Our computational

model predicts that the friction scales exponentially—rather than linearly—with the number of crosslinking Ase1 molecules (Figure 3E). Our experiments provide three independent lines of evidence for this non-linear behavior: (1) the sliding velocities measured in the absence of crosslinker condensation (Figures 3B–3D), as well as (2) the halting of overlap expansion in the presence of crosslinker condensation (Figure 4) cannot be explained by a linear model, and (3) direct measurements of the friction coefficient as a function of the number of crosslinking Ase1 molecules show an exponential relation (Figure 3F). Ase1 oligomerization (as reported in Kapitein et al., 2008) could potentially result in superlinear scaling, because the movement of the monomers in an oligomer becomes tightly coupled. However, under our experimental conditions we did not observe Ase1 oligomers (e.g., Figures 1D, 4B, 5D, and 5E). Moreover, the Hill coefficient for Ase1 binding was very low (Figure S3A), indicating that binding is essentially non-cooperative and that the energetic interactions between the bound Ase1 molecules were very weak. In our computational model, we therefore assumed that Ase1 binds non-cooperatively to microtubules, and we nonetheless found that the friction increases exponentially with the number of crosslinkers (Figure 3E). We attribute the exponential scaling of the friction to the fact that filament movement is a collective and activated process that requires the “simultaneous” hopping (i.e., transient unbinding) of multiple crosslinkers. This process involves the crossing of an energy barrier that increases linearly with the number of crosslinkers, leading to an exponential decrease in the rate of crossing the barrier (Erickson, 2009; Volkov et al., 2013) (Extended Results, Text 4). We hypothesize that this mechanism might be relevant also for other processes where cellular structures are tethered to microtubules. For example, fewer microtubule-interacting proteins than expected based on a linear dependence of the friction may be sufficient to forcefully tether kinetochores to microtubules.

### Entropic Forces Are Generated whenever Molecular Diffusion Is Confined

The mechanism of entropic force generation by confined molecules is a universal phenomenon beyond the Ase1/PRC1/MAP65 family of microtubule crosslinking proteins. Recently, nucleosome unwrapping was quantitatively explained by the one-dimensional pressure exerted by DNA binding proteins diffusing along a DNA strand (Forties et al., 2011). Furthermore, entropic forces are also generated in 2D systems, which is exemplified by the finding that crowding of membrane-bound proteins generates a lateral pressure, which can bend membranes (Stachowiak et al., 2012). Concerning the cytoskeleton, it has long been believed that the constriction of the actin contractile ring is driven by non-muscle myosin II (NMII) translocation of actin filaments. However, recent experiments indicate that NMII is required not for its motor activity to translocate actin, but for its capacity to crosslink actin filaments (Ma et al., 2012). Our results suggest that NMII may be able to generate tension between actin filaments via the mechanism of entropic expansion if it can diffuse between filaments.

Our in vitro system allows for the well-controlled experimental investigation of the interplay between entropic-expansion

forces, crosslinker-condensation forces, and crosslinker-frictional forces that drive the sliding of filaments relative to each other. By examining a minimal system consisting of crosslinkers and microtubules, outside of the cytoplasm, we gain access to biophysical properties of the system that are impossible to access in vivo where they are obscured by numerous interdependent processes. Taken together, our results demonstrate that the thermal motion of confined crosslinkers constitutes a force-producing element within self-organizing filamentous networks, which can complement forces generated by molecular motors and filament dynamics.

## EXPERIMENTAL PROCEDURES

### Protein Purification

Recombinant histidine-tagged full-length *S. pombe* Ase1-GFP (Figure S1A) and *D. melanogaster* Ncd and GFP-Ncd were expressed and purified as described previously (Fink et al., 2009; Janson et al., 2007).

### Sample Preparation

Microtubules and flow chambers were prepared as described previously (Fink et al., 2009). If not noted otherwise, dimly rhodamine-labeled, biotinylated template microtubules in BRB80 buffer (80 millimolar (mM) PIPES, 1 mM EGTA, 1 mM  $MgCl_2$ , 10  $\mu$ M paclitaxel, pH 6.9) were injected into the flow chamber and bound in an aligned manner to surface-immobilized biotin anti-bodies (Sigma, B3640). After rinsing the chamber with assay buffer (20 mM HEPES at pH 7.2, 1 mM EGTA, 0.1 mM EDTA, 75 mM KCl, 1 mM ATP (+Mg), 10 mM DTT, 0.5 mg/ml casein, 10  $\mu$ M paclitaxel, 0.1% Tween, 20 mM D-glucose, 110  $\mu$ g/ml glucose oxidase, and 20  $\mu$ g/ml catalase), 50 pM Ase1-GFP was flushed in, which bound to the template microtubules. In the next step, brightly rhodamine-labeled transport microtubules were flushed in (without Ase1-GFP in that solution) and allowed to bind to the template microtubules that were still covered sparsely with Ase1-GFP. For hydrodynamic flow experiments, finally a 30 s long steady flow of assay buffer (without Ase1-GFP) was applied to shorten the microtubule overlaps by sliding the transport microtubules along the template microtubules, while concurrently removing all unbound transport microtubules. For Ase1-condensation experiments the duration of the final step was 5 s and the buffer included 17 pM Ase1-GFP. For Ncd-Ase1-sliding experiments, the duration of the final step was 5 s and the buffer included 312 pM Ase1-GFP and 300 pM Ncd. For microtubule-microtubule diffusion experiments the template microtubules were Cy-5 labeled in order to allow for high-precision position tracking of the brightly rhodamine-labeled transport microtubules (since tracking accuracy would be impaired if transport and template microtubules had the same fluorescent label) and the duration of the final step was approximately 5 s. For optical trapping experiments dimly Cy5-labeled, digoxigeninated template microtubules in BRB80 buffer were bound to surface-immobilized digoxigenin anti-bodies (Roche, # 11333089001). After rinsing the chamber with assay buffer, Ase1-GFP was flushed in and brightly Cy5-labeled, biotinylated transport microtubules were subsequently flushed in (no Ase1-GFP in solution, neither here, nor in the following steps). In the next step, assay buffer with NeutrAvidin coated silica microspheres was applied. Using a trapped microsphere attached to a biotinylated transport microtubule, overlaps were shortened by moving the template microtubule with a piezo stage.

### Image Acquisition during Hydrodynamic Flow, Microtubule-Microtubule Diffusion, Ase1-Condensation, and Ncd-Ase1-Sliding Experiments

Rhodamine-labeled microtubules, Cy-5 labeled microtubules, and Ase1-GFP were visualized sequentially by switching between tetramethylrhodamine isothiocyanate (TRITC), Cy-5, and GFP filters (Chroma Technology), respectively, using a previously described setup (Fink et al., 2009) with acquisition rates of one frame per 6 or 30 s (time-lapse information indicated in the figures).



### Image Analysis of Microtubule-Microtubule Sliding and Diffusion Experiments

In the hydrodynamic flow, Ase1-condensation and Ncd-sliding experiments the positions of the transport microtubules relative to the template microtubules were determined in each frame. Partial microtubule-overlaps had a non-moving boundary (corresponding to the end of the template microtubule, which is fixed on the coverslip) and a moving boundary (corresponding to the end of the transport microtubule, which moves along the template). The moving ends were read out from the TRITC channel as the positions of the transport microtubule ends. Using the fact that Ase1-GFP bound more strongly to the overlaps as compared to single microtubules, the non-moving ends were read out from the GFP channel as the positions of the edges of the GFP signals averaged over all frames of a time-lapse movie. Sliding velocities were obtained from positional data of the transport microtubules using a rolling frame average over five frames. In the microtubule-microtubule diffusion experiments, image analysis was performed similarly to the experiments described above with the exception of using a high-precision tracking software, Fiesta, to determine the drift corrected positions of the transport microtubules (Ruhnow et al., 2011). Drift correction was performed in Fiesta by tracking the positions of 200 nanometer (nm) TetraSpeck beads (Life Technologies) non-specifically attached to the coverslip surface.

### Optical Trapping and Analysis

An optical tweezers setup (JPK Instruments, NanoTracker) was built on a Nikon eclipse Ti microscope equipped with a Nikon TIRF 60× N.A. 1.49 objective. Lateral bead positions were inferred by back focal plane detection using a quadrant photo diode. Sensitivity and stiffness were obtained using a built-in calibration feature that fits a Lorentzian function to the power spectrum of the thermal fluctuations of a trapped bead. Carboxylated silica beads (Bangs Laboratories, #SC04N) were functionalized with NeutrAvidin (ThermoScientific) using 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide-N-hydroxysulfosuccinimide chemistry. All measurements were performed at a trap stiffness of approximately 0.15 pN/nm (750 milliwatt optical power of a 1,064 nm infrared-laser), and the time traces were recorded with 5 kHz sampling rate. The time traces were converted from voltages to forces and further analyzed using MATLAB. A constant offset, given by the averaged signal after microtubule separation, was subtracted from all forces. To estimate the Ase1-generated entropic forces, we averaged the detected forces starting after 2 s relaxation time after each rapid movement of the piezo stage (see Figure S2B). In Figure 2C, we only present data that were recorded after the bead was pulled past the end of the template microtubule, such that the forces applied to the overlap were solely pulling forces. Negative forces with amplitudes smaller than 0.5 pN were occasionally observed in individual traces due to drift in the optical tweezers setup. For optical imaging, the tweezers setup was equipped with a Nikon TIRF microscopy unit, which was used to visualize the microtubule overlaps. Cy5-labeled microtubules and Ase1-GFP were excited sequentially using 642 nm and 488 nm lasers (Vortran) and a dual-band filter set (Chroma Technology). Image acquisition was performed by a back illuminated EMCCD camera (Andor) at rates of one frame per 10 s (time-lapse information indicated in the figures) using Micro-Manager software (Edelstein et al., 2010).

### Estimating the Number of GFP-Ase1 Molecules in Microtubule Overlaps

The location of a microtubule overlap (either determined by the enhanced GFP-signal or the positions of the template and transport microtubules) was used as mask to read out the integrated Ase1-GFP signal in an overlap. The fluorescence signal (obtained with the same filter set) integrated over the same mask area directly adjacent to the overlap was subtracted as the background signal. The “Ase1-GFP fluorescence intensity” in a microtubule overlap (as used in Figures 2C inset, 3F, 4C, and 5B) was then calculated by dividing the background-corrected integrated Ase1-GFP signal by the fluorescence signal of a single Ase1-GFP molecule, as described previously (Braun et al., 2011). The Ase1-GFP fluorescence intensity thus provides a rough estimate of the absolute number of GFP-Ase1 molecules in a microtubule overlap and can, most importantly, be used to study entropic force, friction, and overlap length as function of relative changes in the number of GFP-Ase1 molecules in an overlap. However, due to significant errors inherently associated

with the described procedure (e.g., experimental uncertainties due to GFP bleaching and blinking, as well as the uneven TIRF illumination, the extent of which may vary from experiment to experiment), we refrain from equating the Ase1-GFP fluorescence intensity with the actual number of GFP-Ase1 molecules in a microtubule overlap and rather express it in AU.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, five figures, one table, and five movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.051>.

### AUTHOR CONTRIBUTIONS

Experiments were conceived and analyzed by Z.L., M.B., A.L., P.R.t.W., M.E.J., and S.D. and performed by Z.L., M.B., and A.L. P.R.t.W. developed the theory and supervised the mathematical modeling; A.L. and M.S. optimized the optical tweezers setup; and Z.L., M.B., A.L., P.R.t.W., M.E.J., and S.D. discussed the results and wrote the manuscript.

### ACKNOWLEDGMENTS

We thank Stephan Grill, Joe Howard, Tim Mitchison, François Nédélec, Friedrich Schwarz, Antoine van Oijen, and Klaus Kroy for discussions, as well as Gijssje Koenderink and Marileen Dogterom for comments on an earlier version of the manuscript. M.B., Z.L., A.L., and S.D. acknowledge support from the European Research Council (ERC starting grant 242933 to S.D.), the Deutsche Forschungsgemeinschaft (Heisenberg programme grant DI 1226/4 and research unit SFG 877 grant DI 1226/4), and the Dresden International Graduate School for Biomedicine and Bioengineering (stipend to A.L.); M.S. from the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung grant 03Z2EN11); and M.E.J., and P.R.t.W. from the Foundation for Fundamental Research on Matter (FOM) and the Division for Earth and Life Sciences (ALW grant 834.09.005), which are part of the Netherlands Organisation for Scientific Research (NWO).

Received: July 21, 2014

Revised: November 11, 2014

Accepted: January 6, 2015

Published: March 5, 2015

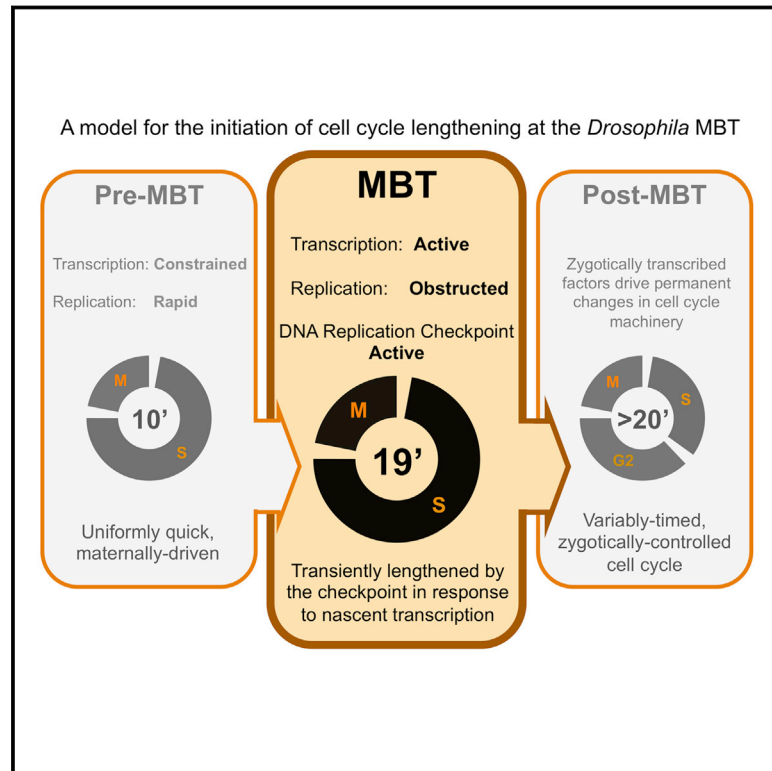
### REFERENCES

- Asbury, C.L., Gestaut, D.R., Powers, A.F., Franck, A.D., and Davis, T.N. (2006). The Dam1 kinetochore complex harnesses microtubule dynamics to produce force and movement. *Proc. Natl. Acad. Sci. USA* 103, 9873–9878.
- Bieling, P., Telley, I.A., and Surrey, T. (2010). A minimal midzone protein module controls formation and length of antiparallel microtubule overlaps. *Cell* 142, 420–432.
- Bormuth, V., Varga, V., Howard, J., and Schäffer, E. (2009). Protein friction limits diffusive and directed movements of kinesin motors on microtubules. *Science* 325, 870–873.
- Braun, M., Drummond, D.R., Cross, R.A., and McAnish, A.D. (2009). The kinesin-14 Klp2 organizes microtubules into parallel bundles by an ATP-dependent sorting mechanism. *Nat. Cell Biol.* 11, 724–730.
- Braun, M., Lansky, Z., Fink, G., Ruhnow, F., Diez, S., and Janson, M.E. (2011). Adaptive braking by Ase1 prevents overlapping microtubules from sliding completely apart. *Nat. Cell Biol.* 13, 1259–1264.
- Civelekoglu-Scholey, G., and Scholey, J.M. (2010). Mitotic force generators and chromosome segregation. *Cell. Mol. Life Sci.* 67, 2231–2250.
- Compton, D.A. (1998). Focusing on spindle poles. *J. Cell Sci.* 111, 1477–1481.
- Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer control of microscopes using µManager. *Curr. Protoc. Mol. Biol. Chapter* 14, Unit14.20.

- Einstein, A. (1906). Zur theorie der Brownschen bewegung. *Annalen der Physik* 19, 371–381.
- Erickson, H.P. (2009). Modeling the physics of FtsZ assembly and force generation. *Proc. Natl. Acad. Sci. USA* 106, 9238–9243.
- Fink, G., Hajdo, L., Skowronek, K.J., Reuther, C., Kasprzak, A.A., and Diez, S. (2009). The mitotic kinesin-14 Ncd drives directional microtubule-microtubule sliding. *Nat. Cell Biol.* 11, 717–723.
- Forth, S., Hsia, K.-C., Shimamoto, Y., and Kapoor, T.M. (2014). Asymmetric friction of nonmotor MAPs can lead to their directional motion in active microtubule networks. *Cell* 157, 420–432.
- Forties, R.A., North, J.A., Javadi, S., Tabbaa, O.P., Fishel, R., Poirier, M.G., and Bundschuh, R. (2011). A quantitative model of nucleosome dynamics. *Nucleic Acids Res.* 39, 8306–8313.
- Fu, C., Ward, J.J., Loïdice, I., Velve-Casquillas, G., Nedelec, F.J., and Tran, P.T. (2009). Phospho-regulated interaction between kinesin-6 Klp9p and microtubule bundler Ase1p promotes spindle elongation. *Dev. Cell* 17, 257–267.
- Furuta, K., Furuta, A., Toyoshima, Y.Y., Amino, M., Oiwa, K., and Kojima, H. (2013). Measuring collective transport by defined numbers of processive and nonprocessive kinesin motors. *Proc. Natl. Acad. Sci. USA* 110, 501–506.
- Gayathri, P., Fujii, T., Møller-Jensen, J., van den Ent, F., Namba, K., and Löwe, J. (2012). A bipolar spindle of antiparallel ParM filaments drives bacterial plasmid segregation. *Science* 338, 1334–1337.
- Gestaut, D.R., Graczyk, B., Cooper, J., Widlund, P.O., Zelter, A., Wordeman, L., Asbury, C.L., and Davis, T.N. (2008). Phosphoregulation and depolymerization-driven movement of the Dam1 complex do not require ring formation. *Nat. Cell Biol.* 10, 407–414.
- Helenius, J., Brouhard, G., Kalaidzidis, Y., Diez, S., and Howard, J. (2006). The depolymerizing kinesin MCAK uses lattice diffusion to rapidly target microtubule ends. *Nature* 441, 115–119.
- Hentrich, C., and Surrey, T. (2010). Microtubule organization by the antagonistic mitotic motors kinesin-5 and kinesin-14. *J. Cell Biol.* 189, 465–480.
- Hill, T.L. (1985). Theoretical problems related to the attachment of microtubules to kinetochores. *Proc. Natl. Acad. Sci. USA* 82, 4404–4408.
- Hunt, A.J., Gittes, F., and Howard, J. (1994). The force exerted by a single kinesin molecule against a viscous load. *Biophys. J.* 67, 766–781.
- Janson, M.E., Loughlin, R., Loïdice, I., Fu, C., Brunner, D., Nédélec, F.J., and Tran, P.T. (2007). Crosslinkers and motors organize dynamic microtubules to form stable bipolar arrays in fission yeast. *Cell* 128, 357–368.
- Kapitein, L.C., Peterman, E.J.G., Kwok, B.H., Kim, J.H., Kapoor, T.M., and Schmidt, C.F. (2005). The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks. *Nature* 435, 114–118.
- Kapitein, L.C., Janson, M.E., van den Wildenberg, S.M.J.L., Hoogenraad, C.C., Schmidt, C.F., and Peterman, E.J.G. (2008). Microtubule-driven multimerization recruits ase1p onto overlapping microtubules. *Curr. Biol.* 18, 1713–1717.
- Lan, G., Daniels, B.R., Dobrowsky, T.M., Wirtz, D., and Sun, S.X. (2009). Condensation of FtsZ filaments can drive bacterial cell division. *Proc. Natl. Acad. Sci. USA* 106, 121–126.
- Loïdice, I., Staub, J., Setty, T.G., Nguyen, N.P., Paoletti, A., and Tran, P.T. (2005). Ase1p organizes antiparallel microtubule arrays during interphase and mitosis in fission yeast. *Mol. Biol. Cell* 16, 1756–1768.
- Ma, X., Kovács, M., Conti, M.A., Wang, A., Zhang, Y., Sellers, J.R., and Adelstein, R.S. (2012). Nonmuscle myosin II exerts tension but does not translocate actin in vertebrate cytokinesis. *Proc. Natl. Acad. Sci. USA* 109, 4509–4514.
- Nelson, S.R., Trybus, K.M., and Warshaw, D.M. (2014). Motor coupling through lipid membranes enhances transport velocities for ensembles of myosin Va. *Proc. Natl. Acad. Sci. USA* 111, E3986–E3995.
- Neujahr, R., Heizer, C., and Gerisch, G. (1997). Myosin II-independent processes in mitotic cells of Dictyostelium discoideum: redistribution of the nuclei, re-arrangement of the actin system and formation of the cleavage furrow. *J. Cell Sci.* 110, 123–137.
- Peskin, C.S., Odell, G.M., and Oster, G.F. (1993). Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J.* 65, 316–324.
- Peterman, E.J.G., and Scholey, J.M. (2009). Mitotic microtubule crosslinkers: insights from mechanistic studies. *Curr. Biol.* 19, R1089–R1094.
- Powers, A.F., Franck, A.D., Gestaut, D.R., Cooper, J., Graczyk, B., Wei, R.R., Wordeman, L., Davis, T.N., and Asbury, C.L. (2009). The Ndc80 kinetochore complex forms load-bearing attachments to dynamic microtubule tips via biased diffusion. *Cell* 136, 865–875.
- Roostalu, J., Hentrich, C., Bieling, P., Telley, I.A., Schiebel, E., and Surrey, T. (2011). Directional switching of the kinesin Cin8 through motor coupling. *Science* 332, 94–99.
- Ruhnow, F., Zwicker, D., and Diez, S. (2011). Tracking single particles and elongated filaments with nanometer precision. *Biophys. J.* 100, 2820–2828.
- Schuyler, S.C., Liu, J.Y., and Pellman, D. (2003). The molecular function of Ase1p: evidence for a MAP-dependent midzone-specific spindle matrix. Microtubule-associated proteins. *J. Cell Biol.* 160, 517–528.
- Stachowiak, J.C., Schmid, E.M., Ryan, C.J., Ann, H.S., Sasaki, D.Y., Sherman, M.B., Geissler, P.L., Fletcher, D.A., and Hayden, C.C. (2012). Membrane bending by protein-protein crowding. *Nat. Cell Biol.* 14, 944–949.
- Subramanian, R., Wilson-Kubalek, E.M., Arthur, C.P., Bick, M.J., Campbell, E.A., Darst, S.A., Milligan, R.A., and Kapoor, T.M. (2010). Insights into antiparallel microtubule crosslinking by PRC1, a conserved nonmotor microtubule binding protein. *Cell* 142, 433–443.
- Sun, S.X., Walcott, S., and Wolgemuth, C.W. (2010). Cytoskeletal cross-linking and bundling in motor-independent contraction. *Curr. Biol.* 20, R649–R654.
- Tawada, K., and Sekimoto, K. (1991). Protein friction exerted by motor enzymes through a weak-binding interaction. *J. Theor. Biol.* 150, 193–200.
- Valentine, M.T., Fordyce, P.M., Krzysiak, T.C., Gilbert, S.P., and Block, S.M. (2006). Individual dimers of the mitotic kinesin motor Eg5 step processively and support substantial loads in vitro. *Nat. Cell Biol.* 8, 470–476.
- Volkov, V.A., Zaytsev, A.V., Gudimchuk, N., Grissom, P.M., Gintsburg, A.L., Ataulkhanov, F.I., McIntosh, J.R., and Grishchuk, E.L. (2013). Long tethers provide high-force coupling of the Dam1 ring to shortening microtubules. *Proc. Natl. Acad. Sci. USA* 110, 7708–7713.
- Yamashita, A., Sato, M., Fujita, A., Yamamoto, M., and Toda, T. (2005). The roles of fission yeast ase1 in mitotic cell division, meiotic nuclear oscillation, and cytokinesis checkpoint signaling. *Mol. Biol. Cell* 16, 1378–1395.
- Zandi, R., Reguera, D., Rudnick, J., and Gelbart, W.M. (2003). What drives the translocation of stiff chains? *Proc. Natl. Acad. Sci. USA* 100, 8649–8653.

# Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition

## Graphical Abstract



## Authors

Shelby A. Blythe, Eric F. Wieschaus

## Correspondence

efw@princeton.edu

## In Brief

The maternal-to-zygotic transition during early development results in a conflict between de novo RNA polymerase recruitment and ongoing DNA replication. The replication checkpoint functions in this context as a feedback mechanism to drive the initial steps of cell-cycle remodeling in response to increased cellular transcription.

## Highlights

- Dosage of transcribed DNA correlates with degree of replication checkpoint activity
- Time-resolved RNA Pol II ChIP-seq over the course of zygotic genome activation
- Sites of stalled DNA replication overlap with transcribed genomic loci
- Reducing zygotic transcription eliminates a genetic requirement for the checkpoint

## Accession Numbers

GSE62925



# Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition

Shelby A. Blythe<sup>1</sup> and Eric F. Wieschaus<sup>1,\*</sup>

<sup>1</sup>Department of Molecular Biology, Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA

\*Correspondence: [efw@princeton.edu](mailto:efw@princeton.edu)

<http://dx.doi.org/10.1016/j.cell.2015.01.050>

## SUMMARY

A conserved feature of the midblastula transition (MBT) is a requirement for a functional DNA replication checkpoint to coordinate cell-cycle remodeling and zygotic genome activation (ZGA). We have investigated what triggers this checkpoint during *Drosophila* embryogenesis. We find that the magnitude of the checkpoint scales with the quantity of transcriptionally engaged DNA. Measuring RNA polymerase II (Pol II) binding at 20 min intervals over the course of ZGA reveals that the checkpoint coincides with widespread de novo recruitment of Pol II that precedes and does not require a functional checkpoint. This recruitment drives slowing or stalling of DNA replication at transcriptionally engaged loci. Reducing Pol II recruitment in *zelda* mutants both reduces replication stalling and bypasses the requirement for a functional checkpoint. This suggests a model where the checkpoint functions as a feedback mechanism to remodel the cell cycle in response to nascent ZGA.

## INTRODUCTION

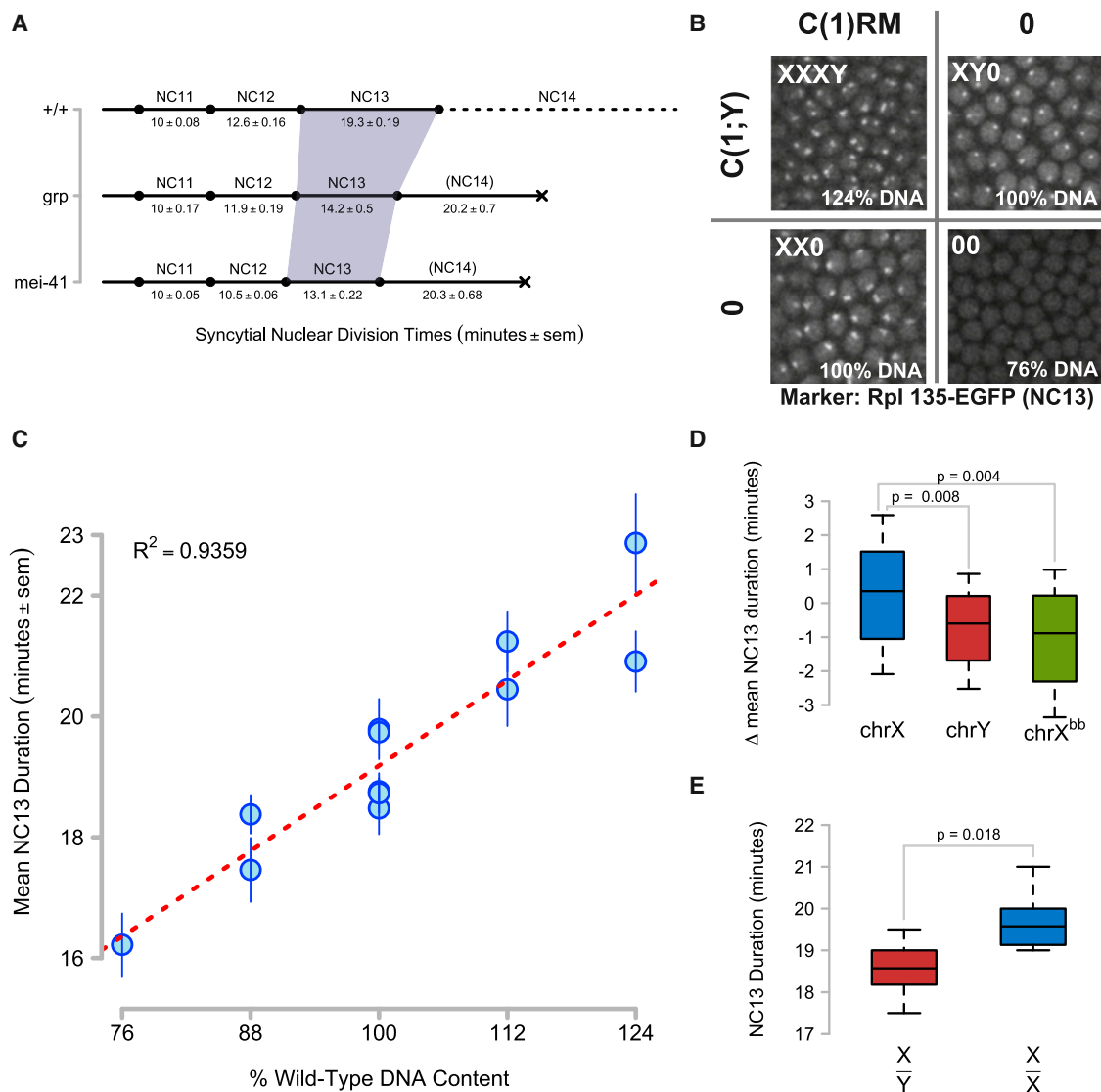
Embryogenesis initiates with a period of cellular proliferation with minimal changes in cellular differentiation and functional specialization (O'Farrell et al., 2004; Tadros and Lipshitz, 2009). In embryos from *Drosophila*, *Xenopus*, and Zebrafish, cellular proliferation occurs with an abbreviated cell cycle consisting of sequential S and M phases without intervening gap phases. Concurrently, constrained transcriptional activity suppresses zygotic patterning in response to maternal cell fate determinants. Upon reaching a precise nucleocytoplasmic (N:C) ratio, embryos undergo coordinated cell-cycle remodeling and large-scale zygotic gene activation (ZGA) and enter a period of cell fate specification and morphogenesis with reduced cellular proliferation. This remodeling of cell-cycle behavior and transcription accompanies a genetic transition from maternal to zygotic control of development collectively termed the midblastula transition (MBT). Although the temporal control of MBT timing via the N:C ratio is precise and reproducible within species, little is

known about how the nuclear content is measured and how the resultant N:C ratio regulates the cell cycle and ZGA.

One attractive candidate for a “sensor” is the DNA replication checkpoint, whose activity is necessary for cell-cycle remodeling and for maintaining ZGA (Brodsky et al., 2000; Conn et al., 2004; Crest et al., 2007; Di Talia et al., 2013; Fogarty et al., 1994; Sibon et al., 1999; Sibon et al., 1997). In *Drosophila* embryos, for example, rapid early mitoses are followed by a gradual checkpoint-mediated lengthening of the final pre-MBT cell cycles. The effect of this checkpoint is most obvious in *Drosophila* at nuclear cycle 13 (NC13), when it is required to extend interphase from 12 to 19 min. *Drosophila* mutants for two checkpoint kinases, ataxia telangiectasia and Rad3-related (*mei-41/ATR*) and checkpoint kinase 1 (*grapes/chk1*), fail to trigger a checkpoint at NC13 and prematurely enter mitosis prior to completion of S phase, resulting in catastrophic DNA damage and ultimately death (Fogarty et al., 1994; Sibon et al., 1997, 1999). The cues that activate the DNA damage response at the MBT are not known. It has been proposed that replication factors become limiting as embryos approach the N:C ratio, thus causing replication stress and triggering the checkpoint (Dasso and Newport, 1990; Sibon et al., 1999). Some support for this model comes from *Xenopus* embryos, where overexpression of a subset of replication factors will increase the number of pre-MBT mitoses from 12–14 to 13–15 (Collart et al., 2013).

To understand further the workings of the MBT clock in *Drosophila*, we have investigated the molecular mechanism that activates the MBT replication checkpoint. Rather than focusing on additional mitoses and other post-MBT (NC14) phenotypes associated with checkpoint defects, we directly test how checkpoint activity scales with the dosage of zygotic DNA at NC13. We find a non-equivalence of genomic DNA for triggering the checkpoint that correlates with the relative quantity of transcriptionally engaged DNA. By use of time-resolved chromatin immunoprecipitation sequencing (ChIP-seq) analysis of RNA Pol II occupancy over the course of ZGA, we determine that checkpoint activation at NC13 likewise correlates with the induction of large-scale de novo binding of Pol II to thousands of promoters. We find evidence that DNA replication slowing or stalling at NC13 co-localizes with and depends upon RNA Pol II activity. Pol II is recruited to chromatin normally in *mei-41* mutant embryos, and reducing Pol II occupancy suppresses the *mei-41* mitotic catastrophe. Thus, we propose that the primary effector downstream of the N:C ratio for timing the MBT is the initial establishment of transcriptional competence at the onset of large-scale ZGA.





**Figure 1. Non-equivalence of Zygotic DNA for MBT Checkpoint Activation**

(A) Timelines of syncytial cell-cycle times for wild-type (+/+), *grp*<sup>1</sup>, and *mei-41*<sup>D3/29D</sup> embryos were measured by time-lapse confocal microscopy of H2Av-GFP or RFP. The shaded region highlights NC13. Lethality is signified by a black X.

(B) Representative confocal images (2,500 μm<sup>2</sup>) of nucleolar RNA Pol I GFP expression in NC13 embryos produced from a cross between C(1)RM/0;Rpl135-EGFP/+ and C(1;Y)1/0 adults. First chromosome dosage is indicated in the upper left of each panel, and the corresponding amount of zygotic genomic DNA is indicated in the bottom left. NC13 nucleolar morphology in XY0 embryos is punctate, whereas it is barbell-shaped in XX0 embryos (See Supplemental Information). No nucleolar Rpl135 EGFP is detected at NC13 in 00 embryos.

(C) NC13 times were measured for embryos with zygotic DNA dosage between 76% and 124% (see [Experimental Procedures](#)). Mean NC13 times ± SEM for N ≥ 11 embryos per genotype are plotted as a function of zygotic genomic DNA content. Linear regression is represented as a red line.

(D) Box plots showing deviations from mean NC13 time for genotypes differing in chrX (n = 74), chrY (n = 40), or rDNA dosage (X<sup>bb</sup>, n = 41). Brackets indicate the results of two-tailed t tests.

(E) NC13 times for male (X/Y, n = 11) and female (X/X, n = 12) embryos produced from *w*; *His2Av-RFP* × *w*; *HbP2* > *GFPnls* /Y; + adults. Box plots show the distribution of NC13 times for each genotype. Brackets indicate the results of a two-tailed t test.

See also [Figure S1](#).

## RESULTS

### Non-equivalence of Genomic DNA for Triggering the MBT Replication Checkpoint

Following fertilization, *Drosophila* embryos undergo 13 rapid metasynchronous syncytial mitoses, gradually lengthening the

cell-cycle period from an initial period of 8 min prior to nuclear cycle 10 (NC10) to ~19 min at NC13 (Foe and Alberts, 1983) (Figure 1A). The characteristic lengthening of NC13 corresponds to a lengthening of S phase (Shermoen et al., 2010) and therefore serves as a read-out for the magnitude of an induced DNA replication checkpoint. Compared with a 12.6 ± 0.16 min wild-type

NC12, the NC13 is lengthened in wild-type embryos by  $53\% \pm 3\%$ , whereas NC13 is only  $4\% \pm 3\%$  longer in *mei-41* and  $12.7\% \pm 5\%$  longer in *grp* (Figure 1A). This genetic requirement for a replication checkpoint indicates that NC13 embryos may be subject to a new source of replication stress. Unlike in other organisms, this replication stress does not seem to be related directly to replication capacity (Collart et al., 2013; Dasso and Newport, 1990; Sibon et al., 1999), as reducing levels of the 180 kDa subunit of DNA Polymerase  $\alpha$  (Brodsky et al., 2000; LaRocque et al., 2007), or the noncatalytic subunit of Cdc7 kinase *Dbf4/chiffon* has as no impact on NC13 duration (Figure S1).

To test whether the checkpoint scales with the N:C ratio, we measured the correlation of NC13 time with the overall quantity of zygotic genomic DNA. We generated embryos containing between 76% and 124% DNA content by varying the dosage of chromosomes X and Y (chrX, chrY) using compound chromosome stocks (see Experimental Procedures and Figure 1B). The duration of NC13 positively correlates with zygotic DNA content (Figure 1C), but we observe two notable and informative discrepancies. First, the mean duration of NC13 for chrX<sup>+</sup> genotypes is  $1.1 \pm 0.5$  min longer than chrY<sup>+</sup> genotypes of otherwise equivalent DNA content (Figure 1D). We observe a similar discrepancy between male and female embryos in an otherwise wild-type *His2Av-RFP* stock, where X/Y embryos complete NC13 in  $18.7 \pm 0.33$  min whereas X/X embryos complete NC13 in  $19.7 \pm 0.20$  min, a difference of  $1.0 \pm 0.5$  min (Figure 1E). The character of the chrX DNA also influences the mean duration of NC13. Embryos with a wild-type X have NC13 duration that is  $1.3 \pm 0.6$  min longer than those with an X lacking rDNA repeats (Figure 1D). Although small, these differences are significant. If NC13 duration depended solely on absolute DNA content, based on the linear fit of NC13 times to DNA content (Figure 1C, red dashed line), shortening the cycle by 1 min would require reducing DNA dosage by 8.3%, or  $\sim 70\%$  of the first chromosome. We conclude that not all DNA sequences are equivalent for triggering the replication checkpoint at the MBT. One major difference between chrX and chrY is the degree of transcriptionally active tracts of euchromatic DNA, with the X consisting of  $\sim 50\%$  euchromatin and 50% heterochromatin in contrast to the 100% heterochromatic Y. In addition, highly transcribed rDNA repeats also modulate the magnitude of the checkpoint (Figure 1D). Therefore, we set out to test the alternative model that the checkpoint scales with the degree of transcriptionally engaged genomic DNA.

### Large-Scale Recruitment of Poised RNA Pol II Distinguishes NC13 from NC12

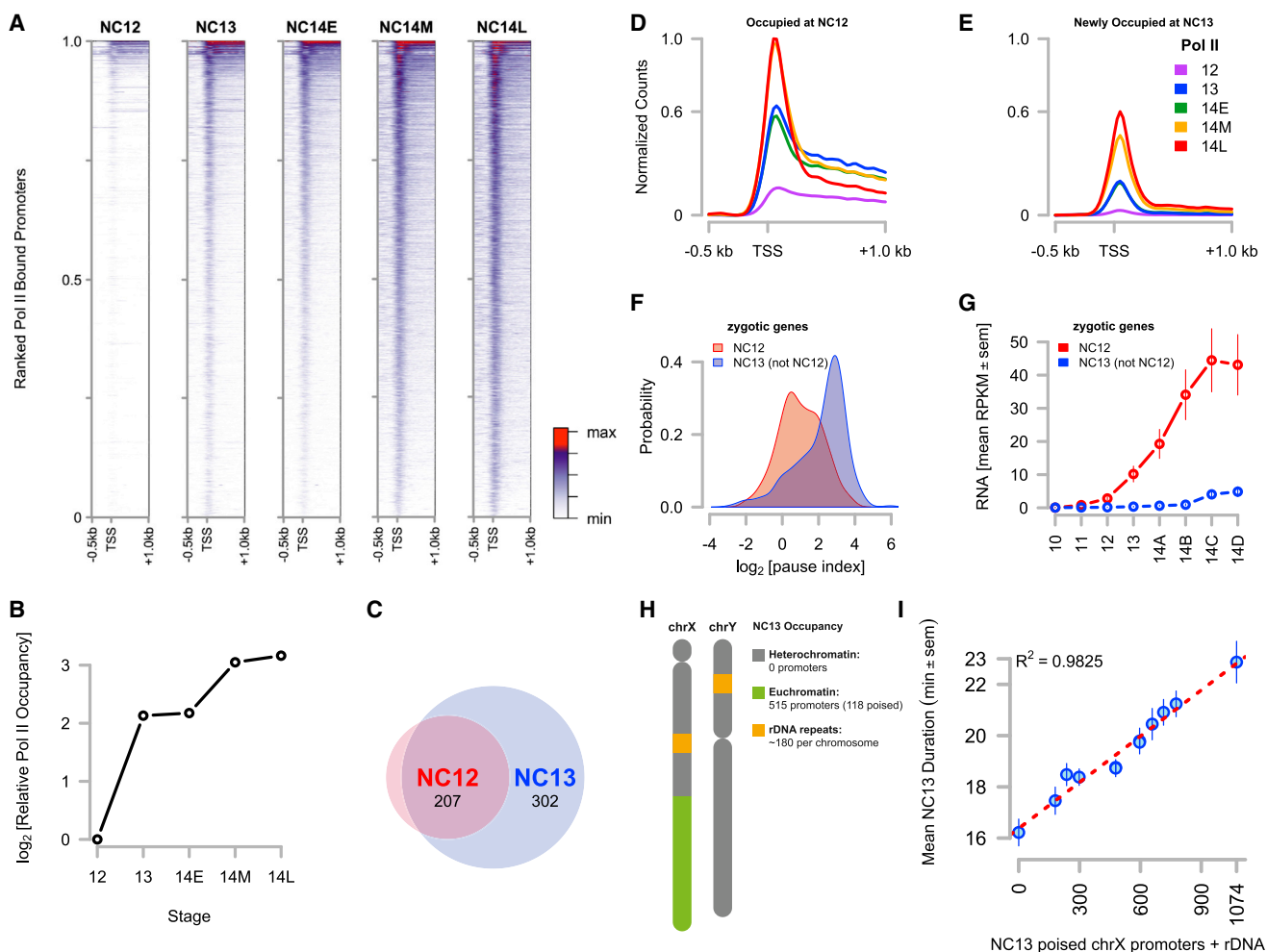
To characterize transcriptional competency in the early embryo, we developed a method for performing ChIP-seq on small numbers of precisely staged embryos to measure the dynamics of RNA Pol II occupancy during ZGA. We carefully optimized sample preparation to generate high-quality measurements of RNA Pol II occupancy using 100–200 embryos collected during single interphases (NC12 or 13), or for three time points within NC14 interphase (Early, Middle, and Late, approximately 0–15', 15–35', and 35–60' after NC14; see Extended Experimental Procedures), with an average interval of 18 min between time points. This approach extends previous analyses of Pol II binding during

ZGA (Chen et al., 2013) allowing the dynamics of RNA Pol II recruitment to be reconstructed at each cycle over the course of the MBT (Figure 2).

At the outset of NC12, a small cohort of pre-MBT transcribed genes is occupied by initiating RNA Pol II (CTD pSer5) ("Pol II" hereafter) (Figures 2A and 2B; Table S1). Between NC12 and NC13, there is a 5.4-fold increase in the number of promoters occupied by Pol II. During NC13, initiating Pol II is significantly enriched within 1 kb of 2,988 promoters, in contrast to NC12, when only 550 promoters are significantly bound. This trend matches the overall quantity of Pol II binding over this time, where between NC12 and NC13, there is a 4.4-fold change in the total Pol II occupancy within the genome (Figure 2B). To characterize this increase in Pol II occupancy between NC12 and NC13, we calculated the mean Pol II distribution over genes within two classes of NC13 peaks: those bound at NC13 and also at NC12, versus those newly occupied at NC13 (Figure 2C). The mean Pol II distribution for genes occupied at NC12 is initially uniformly distributed throughout the gene body, whereas Pol II is largely found concentrated near the transcription start site (TSS) for genes newly bound at NC13 (Figures 2D and 2E).

The distribution of Pol II at genes newly bound at NC13 resembles that of stalled or poised Pol II, previously been shown to be established over the course of the MBT (Chen et al., 2013). To estimate the degree of poising, we calculated a "pause index," in which higher pause indices indicate a greater probability of a gene being poised (Zeitlinger et al., 2007). Indeed, the mean pause index for the set of NC12-bound zygotic genes is significantly different than that of the set of promoters newly bound at NC13 (Figure 2D) (1.00 versus 2.14,  $p < 0.01$ , two-tailed t test). To confirm that Pol II is largely recruited in a poised status at NC13, we extracted RNA expression profiles from a published data set for zygotic genes in each class of promoters (Lott et al., 2011). Of the 550 promoters bound by Pol II at NC12, 233 lack a significant maternal contribution and can be classified as "zygotic only." Similarly, of the 2,988 NC13 promoters, 509 are "zygotic only," and 302 of these are not present in the set of NC12-bound promoters (Figure 2B). Poly-A mRNA expression from the set of NC12-bound zygotic genes begins at or around NC12 and steadily increases over the duration of NC14 (Figure 2C, "NC12"). Little or no new poly-A mRNA expression is detected until late in NC14 for the set of promoters newly bound at NC13 (Figure 2C, "NC13 [not NC12]"), consistent with de novo recruitment of Pol II at NC13 directly into the poised status. We conclude from these experiments that the major qualitative distinction in Pol II characteristics between NC12 and NC13 is the large-scale recruitment of Pol II to previously unbound genes and the subsequent establishment of transcriptional poising.

Importantly, the duration of NC13 correlates better with the number of transcriptionally engaged promoters than with bulk DNA alone. At NC13, Pol II occupies 515 promoters within chrX euchromatin, of which 118 are poised (pause index  $\geq 2$ , Figure 2H). We recovered zero bound promoters on either chrX heterochromatin, or on chrY. In addition, we estimated an average of 180 rDNA repeats per X and Y based on previously published measurements (Long and Dawid, 1980). Re-scaling the x axis of Figure 1E with our estimate of poised chrX



**Figure 2. Large-Scale Recruitment of Poised RNA Pol II at NC13**

(A) Promoter-proximal RNA Pol II (CTD pSer5) was plotted for time points spanning the MBT. Significantly enriched promoters are ranked from the top to the bottom of the y axis by high to low mean intensity over the entire time course. The x axis spans  $-0.5$  kb to  $+1.0$  kb and the TSS is noted. The color bar is at the right hand margin.

(B) The sum of normalized Pol II CPM values for each gene in the *Drosophila* genome was calculated for each time point and plotted as an estimate of total Pol II occupancy over the course of MBT.

(C) The number of purely zygotic genes present at either NC12 or NC13 was determined and plotted as a Venn diagram.

(D and E) Mean distributions of Pol II over promoters occupied at NC12 (D) versus promoters newly occupied at NC13 (E) are plotted per time point. The y axis for both plots represents Pol II counts normalized to the maximum count value in both data sets. The maximum count value for genes newly occupied at NC13 is 0.6 and is noted on both axes.

(F) A kernel density estimate was plotted for the set of pause indices for each gene in both the “bound at NC12” set (red) or the “newly bound at NC13” set (blue).

(G) RPKM values for purely zygotic genes in the “bound at NC12” (red) and the “newly bound at NC13” (blue) sets were extracted from (Lott et al., 2011) and averaged. Mean RPKM values  $\pm$  SEM are plotted from NC10 through NC14.

(H) The schematic representation of chrX and Y showing relative quantities of heterochromatic and euchromatic sequences on each. The observed number of promoters occupied and poised at NC13 is annotated on the right.

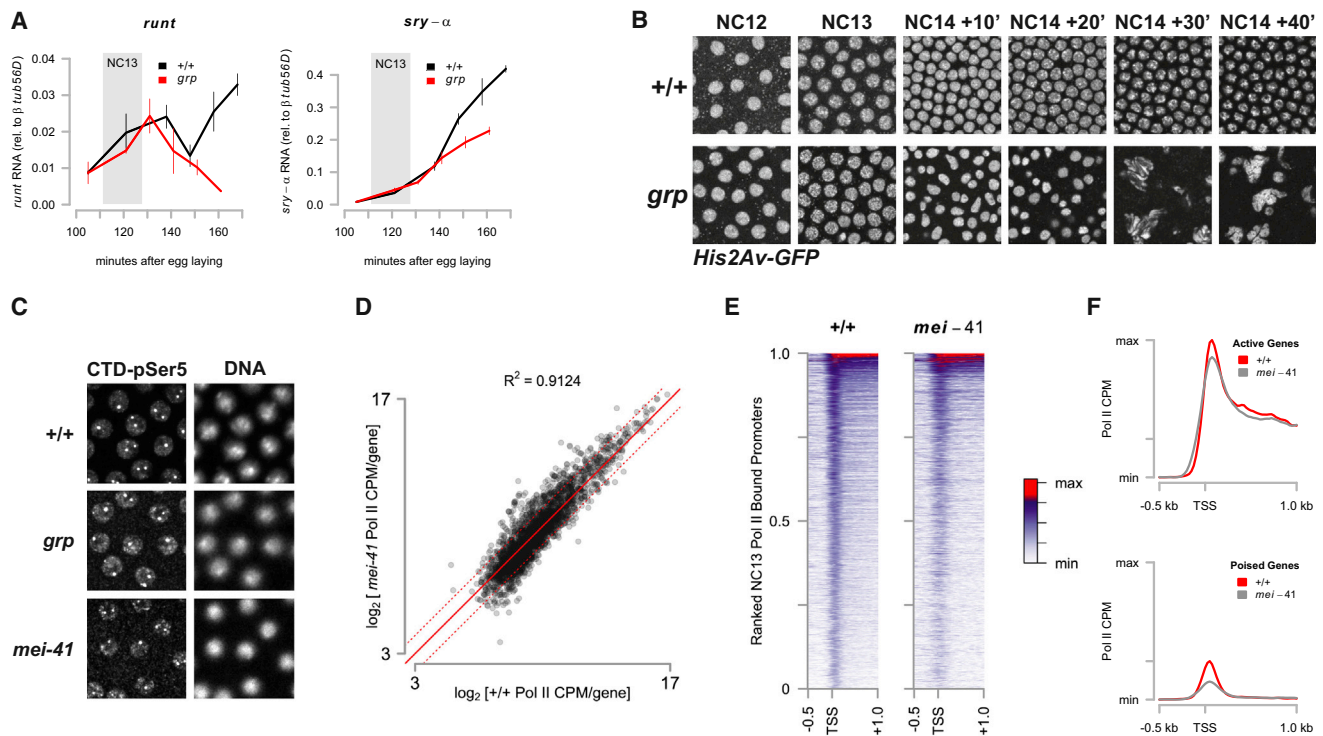
(I) NC13 cell-cycle time data for different X-Y chromosome combinations from Figure 1E is re-scaled and plotted according to the sum of poised chrX promoters plus rDNA repeats. Data are represented as mean  $\pm$  SEM with a linear regression (red line).

See also Table S1.

promoters plus rDNA repeats at NC13 yields a better correlation between the measured NC13 times than DNA content alone (Figure 2I). These results suggest that some property of transcriptionally engaged chromatin architecture presents an unforeseen challenge to the DNA replication machinery, such that its abrupt establishment at NC13 triggers a replication checkpoint.

### A Functional Replication Checkpoint Is Not Necessary for ZGA

We next revisited the question of whether a functional checkpoint is necessary for ZGA. First, we compared the temporal mRNA expression patterns of two zygotic genes in the NC12-bound class whose expression ultimately requires a functional



**Figure 3. A Functional Replication Checkpoint Is Not Necessary for Zygotic Gene Activation**

(A) Quantitative RT-PCR was performed on random-primed cDNA from precisely staged single *w*; His2Av-GFP (*+/+*, black) or *trans*-heterozygous *w*; *grp*<sup>1/209</sup>; His2Av-GFP (*grp*, red) embryos (n = 3 per time point). Mean expression of *runt* or *sry-α* mRNA  $\pm$  SEM is quantified relative to expression of  $\beta$ -tubulin 56D mRNA. The period corresponding to NC13 in wild-type embryos is highlighted (gray box).

(B) Representative time lapse confocal images (2,500  $\mu\text{m}^2$ ) are shown of His2Av-GFP in wild-type (*+/+*, top) and *grp*<sup>1</sup> mutant embryos (bottom) corresponding to the time points in (A).

(C) Staining of RNA Pol II (CTD pSer5) (left) and DNA (DAPI, right) in wild-type (*+/+*, top), *grp*<sup>1</sup> (middle), and *mei-41*<sup>D3/29D</sup> (bottom).

(D)  $\log_2$ [Pol II CPM] values for genes in the set of NC13-bound promoters were plotted for both wild-type and for *mei-41*<sup>D3</sup> NC13 stage embryos. The solid red line indicates no change between samples and the dotted red lines indicate  $\pm 2$ -fold changes.

(E) Promoter-proximal Pol II counts for both wild-type (*+/+*) and *mei-41*<sup>D3</sup> were plotted as in Figure 2A.

(F) Mean promoter-proximal Pol II counts for the set of “active” (upper panel) or “poised” (lower panel) genes in the wild-type (*+/+*) or *mei-41*<sup>D3</sup> datasets are plotted. The y axis is identical between plots and is scaled to the maximal value plotted.

See also Table S2.

checkpoint: *runt* and *sry-α* (Sibon et al., 1997). Precisely staged single-embryo QPCR in wild-type and *grp* mutant embryos shows that eliminating the checkpoint has no measurable effect on expression until 30 min post NC14 (Figure 3A). The reduction in *runt* and *sry-α* mRNA expression corresponds with a precocious catastrophic mitosis in NC14 (Figures 1A and 3B). Both *runt* and *sry-α* initiate normal expression in the absence of the checkpoint. Likewise, by immunostaining for initiating RNA Pol II (CTD pSer5) in syncytial blastoderm stage embryos, no gross difference in Pol II distribution or intensity is observed between wild-type and checkpoint mutant embryos (Figure 3C). These observations indicate that the initial phase of ZGA proceeds normally in checkpoint mutant embryos.

To confirm this observation, we compared the genome-wide distribution of Pol II in NC13-staged wild-type and *mei-41* mutant embryos. The gene-by-gene distribution of Pol II intensities for both wild-type and *mei-41* mutants is highly correlated (Figures 3D and 3E; Table S2). Notably, the TSS-centered peak of Pol II in *mei-41* is broader and more diffuse (Figures 3E and 3F).

Although the mean promoter-proximal distribution of Pol II between wild-type and *mei-41* is largely unchanged in the set of active genes (Figure 3F, top), the peak corresponding to poised Pol II is reduced by 31% relative to wild-type in *mei-41* in the set of poised genes (Figure 3F, bottom). The overall effect of this reduction is small. Summing normalized Pol II counts over all bound genes, genome-wide Pol II occupancy in *mei-41* is 92% of wild-type. This effect of *mei-41* on poised loci could reflect either a feedback mechanism between *mei-41* and poised Pol II, or could stem from the shorter cell-cycle time of NC13 in *mei-41* mutants, if establishment of poising is sensitive to interphase length as is seen with activation of a subset of zygotic loci (Edgar and Schubiger, 1986). We conclude that embryos lacking a functional replication checkpoint initiate this early phase of large-scale ZGA normally.

#### DNA Replication Stalls at Active and Poised Promoters

The initial event sensed by the replication checkpoint machinery is the formation of tracts of single-stranded DNA at sites of replication stress. These exposed sites of single-stranded DNA are



rapidly bound by the conserved Replication Protein A (Rpa) complex that consists of three subunits (RpA-70, Rpa2, and Rpa3 in *Drosophila*), which recruits ATR to sites of stress (Zou and Elledge, 2003). In addition, Rpa also functions as a DNA elongation factor (reviewed in Wold, 1997). Therefore, as an independent approach to studying the MBT replication checkpoint, we developed a fluorescent RpA-70 reporter in order to measure both optically and by ChIP-seq the magnitude and genomic distribution of sites of stalled DNA replication in the NC13 embryo. To test whether this reporter responds to induced replication stress, we depleted dNTPs from wild-type embryos during NC12 (Figure 4B), triggering a temporary Chk1-dependent replication checkpoint (Fasulo et al., 2012) (Figure 4C). Under conditions of induced replication stress, RpA-70 forms intense foci within nuclei that gradually increase in both number and intensity over the course of the lengthened NC12 interphase.

During early syncytial blastoderm interphases (NC10–12), RpA-70 EGFP is uniformly distributed within the nucleoplasm beginning after nuclear envelope formation (Figure 4A). Upon entry into mitotic prophase, RpA-70 is rapidly exported into the cytoplasm and is undetectable on undamaged chromatin throughout mitosis (Figure 4A). The first deviation from this pattern is observed between 9 and 15 min into NC13 when weak and diffuse foci of RpA-70 EGFP are observed in wild-type embryo. These foci do not persist on chromatin by the time nuclei enter mitosis 13 (Figure 4D, +/+). In contrast, *grp* mutant embryos (and *mei-41*, data not shown) form more intense RpA-70 EGFP foci during NC13 (Figure 4D, *grp* Interphase NC13) that are still visible on condensing chromatin when the nuclei enter their premature mitosis (Figure 4D, *grp* Prophase and Metaphase NC13). These results confirm that in wild-type embryos sites of stalled DNA replication arise by NC13 and are resolved prior to mitosis by activation of the replication checkpoint.

To map the genomic distribution of these RpA-70 foci, we performed ChIP-seq for both RpA-70 and Pol II on NC13 RpA-70 EGFP embryos in parallel, and sites of RpA-70 enrichment were calculated as for the Pol II experiments above. We expected that peaks of RpA-70 above background would correspond to sites of stalled DNA replication (see Discussion). Of the 2,804 peaks of significant enrichment of RpA-70 over input DNA, 81% (2,271) of these peaks overlap with those in the set of Pol II peaks, suggesting that the majority of RpA-70 enrichment over background localizes in the vicinity of transcription units (Figure 5A). The gene-by-gene distribution of promoter-proximal Pol II or RpA-70 in the set of overlapping peaks reveals similar distributions of both proteins (Figure 5B). On average, the distribution of RpA-70 near genes is displaced ~100 bp upstream of Pol II peaks, overlapping with TSSs (Figures 5B and 5C). The intensity of RpA-70 is relatively constant over the set of co-enriched promoters, whereas Pol II occupies a wider range of intensities. These results strongly support a correlation between transcriptionally engaged promoters and sites of stalled DNA replication within NC13 chromatin.

### Elimination of Zelda Activity Reduces RpA-70 Binding to Zelda-Dependent Target Genes

The transcription factor *zelda* (*zld*) is necessary for the expression of a broad set of early zygotic genes and is regarded as a

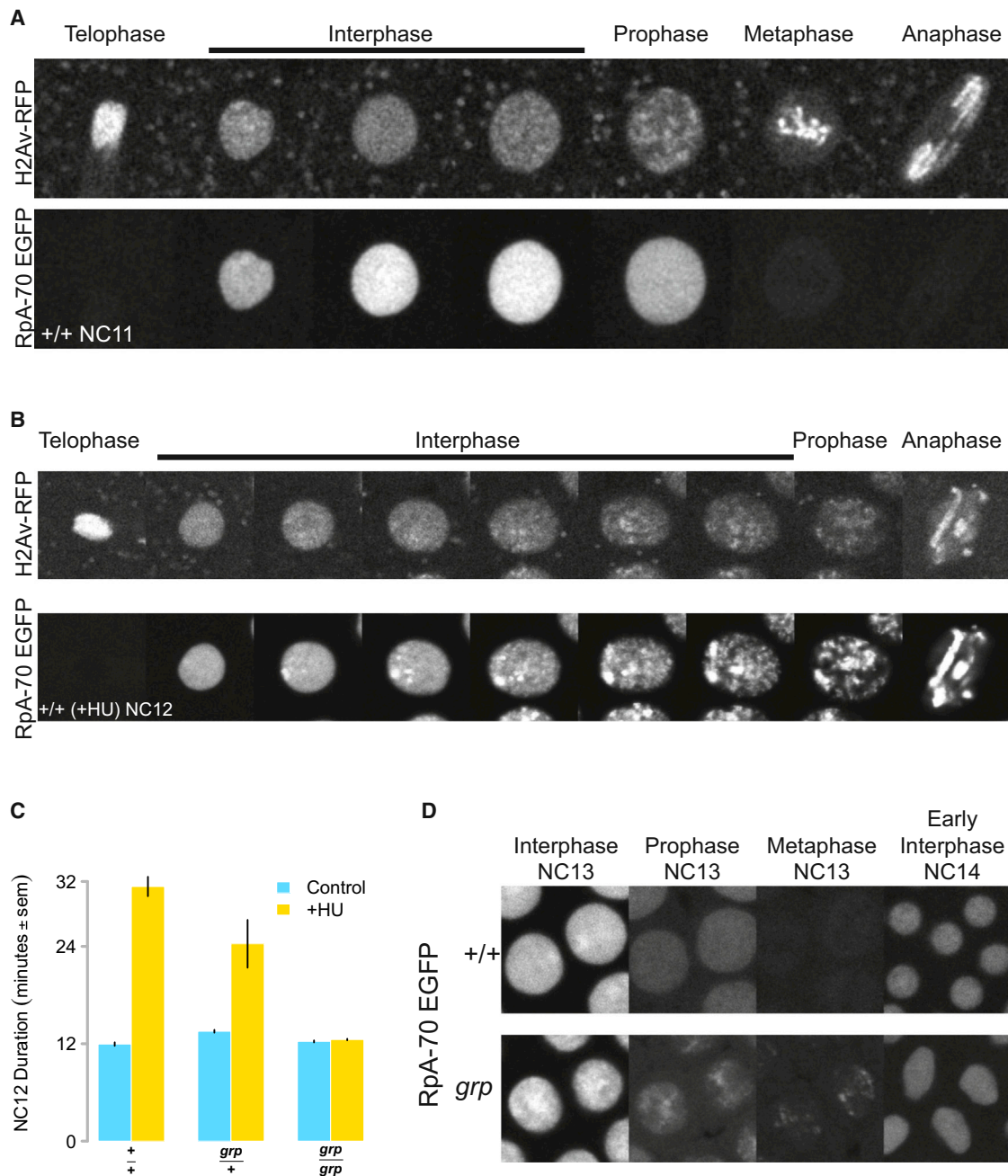
master regulator of ZGA (Harrison et al., 2011; Li et al., 2014; Liang et al., 2008; Nien et al., 2011). We reasoned that Pol II binding to NC13 promoters in *zld* embryos should be largely reduced, and we therefore tested whether reducing Pol II at promoters would have a corresponding effect on the binding of RpA-70.

We performed ChIP-seq for Pol II and RpA-70 as above, comparing wild-type and *zld* mutant embryos. We estimate that overall Pol II binding in *zld* mutant embryos is 51% of wild-type, summing normalized Pol II counts over all NC13-bound genes. The set of genes bound by Pol II in wild-type NC13 embryos can be subdivided into at least two subclasses: *zld*-dependent and *zld*-independent. *zld*-dependent genes were defined as the set of genes where Pol II binding was reproducibly reduced by >2-fold in *zld* embryos at an adjusted p value of <0.01 (Figure 6A; Table S3). The set of *zld*-independent genes was defined as the set of promoters whose Pol II binding was unaffected (i.e., the rounded fold-change between *zld* and wild-type equals zero at an adjusted p value of <0.01) (Figure 6A).

Of the set of 2,988 genes with significant Pol II binding in wild-type NC13 embryos, 435 genes (15%) are *zld*-dependent, of which 266 (61%) fall into the set of genes bound by Pol II at NC12 (Figure 2). In contrast, 1,793 genes (60%) are *zld*-independent, of which 1,671 (93%) are members of the set of genes newly bound at NC13 (Figure 2). *zld* largely affects a subset of the actively expressed genes, consisting of loci with the highest average Pol II count distributions, with little or no effect on the establishment of Pol II binding at the TSS of genes with lower Pol II distributions (Figure 6B), which largely fall into the poised class. Consistent with previous reports (Harrison et al., 2011; Liang et al., 2008), the *zld*-dependent class consists of zygotic genes that are highly expressed early in development (Figure 6C) and are found within  $792 \pm 89$  bp of a Zld binding site (Figures 6D and 6E). In contrast, *zld*-independent genes generally show little or no zygotic expression before the end of NC14 (Figure 6C) and are farther from a mapped Zld binding site ( $5,495 \pm 210$  bp) (Figures 6D and 6E). Correspondingly, *zld* is necessary for Pol II binding to the *zld*-dependent class of genes (Figure 6F), whereas *zld* does not affect Pol II recruitment in the *zld*-independent class (Figure 6F').

We next asked whether reducing Pol II occupancy alters the occupancy of RpA-70 at the *zld*-dependent class of promoters. We compared the distribution of RpA-70 between wild-type and *zld* embryos within *zld*-dependent and -independent promoters (Figures 6G and 6G'). Reduced Pol II occupancy at *zld*-dependent promoters in *zld* mutants corresponds to a reduction in occupancy by RpA-70 (Figure 6G), whereas no change of RpA-70 binding is seen at *zld*-independent promoters (Figure 6G'). This result strongly supports our model that sites of transcriptional activity serve as roadblocks to DNA replication in NC13 embryos.

To test this model, we predicted that reducing total Pol II occupancy at NC13 would suppress the mitotic catastrophe in *mei-41* mutant embryos. Indeed, embryos from *mei-41 zld* double mutant mothers complete the syncytial mitotic divisions without catastrophe in 31% of cases following a short ( $13.8 \pm 0.96$  min) NC13 (Figures 7A and 7B). In contrast, blocking Pol II transcription with  $\alpha$ -amanitin fails to suppress the *mei-41* mitotic catastrophe (Figures 7A and 7B). In the short timescales relevant to this experiment,  $\alpha$ -amanitin functions by inhibiting the



**Figure 4. RPA-70 EGFP Marks Sites of Stalled Replication**

(A) RPA-70 EGFP uniformly localizes to interphase nuclei before NC13. An RPA-70 EGFP; H2Av RFP embryo was imaged by confocal microscopy. Successive representative images of a single NC11 stage nucleus are shown at the cell-cycle stages indicated on top.

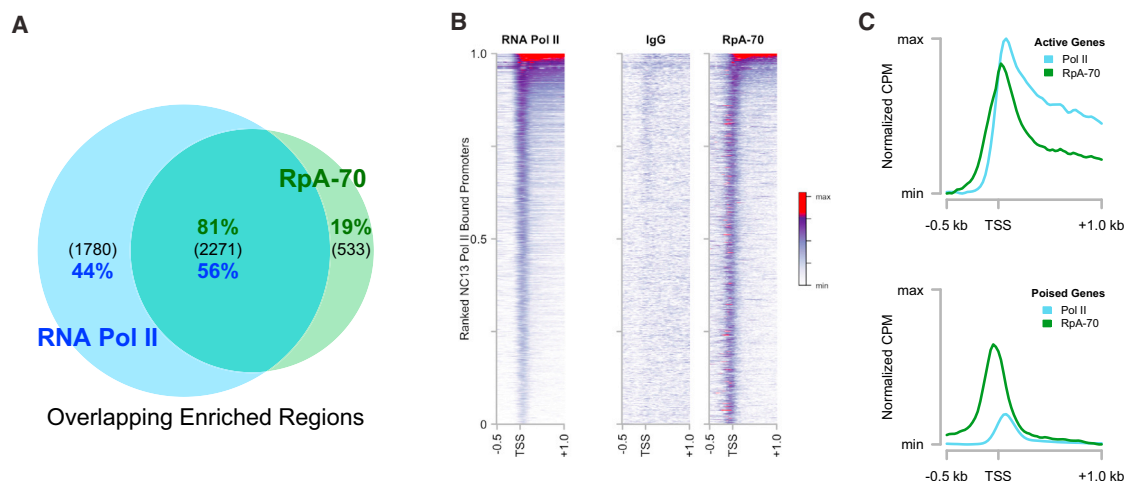
(B) RPA-70 EGFP and H2Av RFP as visualized in a HU-treated embryo by time-lapse confocal microscopy. Successive representative images of a single NC12 nucleus are shown at the cell-cycle stages indicated on top.

(C) Wild-type (*+/+*), *grp* *+*, and *grp* mutant embryos were treated with HU and total NC12 duration was measured by time lapse confocal microscopy.

(D) Wild-type (*+/+*) and *grp* mutant embryos expressing RPA-70 EGFP were visualized by time-lapse confocal microscopy. Successive representative images of two nuclei per genotype are shown at the cell-cycle stages indicated.

translocation of RNA Pol II along DNA (Gong et al., 2004) and does not affect either recruitment of Pol II or initialization of the transcription complex at the TSS (e.g., Li et al., 1996). These re-

sults are consistent with a model in which a feature of ZGA upstream of entry into transcriptional elongation drives DNA replication stalling at NC13.



**Figure 5. Rpa-70 EGFP Co-localizes with RNA Pol II in NC13 Embryos**

(A) Genomic regions significantly enriched by ChIP-seq for Rpa-70 or Pol II were identified and intersected to yield a set of co-enriched regions. The total number of peaks in each group is indicated in parenthesis and the percentage of the total for each ChIP is shown (green, Rpa-70; blue, Pol II). (B and C) Promoter-by-promoter view of Pol II and Rpa-70 localization on NC13 chromatin is shown (B) compared with a control ChIP (IgG). The mean distribution of Pol II and Rpa-70 over actively expressed (C, top) and poised (C, bottom) is plotted. To facilitate comparison, mean CPM values per set of co-enriched promoters for each ChIP were calculated and are presented normalized to the maximal value per ChIP and are then floored to the minimum value (normalized CPM).

Next, we compared *mei-41* suppression by *zld* with heterozygosity for *Cyclin B*, previously reported to suppress MBT replication checkpoint defects (Sibon et al., 1999). Heterozygosity for *Cyclin B* (*Df(CycB)*/+) suppresses *mei-41* in 54% of cases. Unlike *zld*, *CycB*-dependent suppression is accompanied by a significantly lengthened ( $16.8 \pm 1.09$  NC13 time (Figure 7A) (Sibon et al., 1999). Together, these results suggest two mechanisms for suppressing a requirement for a functional DNA replication checkpoint at the MBT, either by reducing the source of replication stalling (e.g., by reducing ZGA via factors such as *zld*), or by providing enough time to complete DNA replication (e.g., by reducing *CycB*).

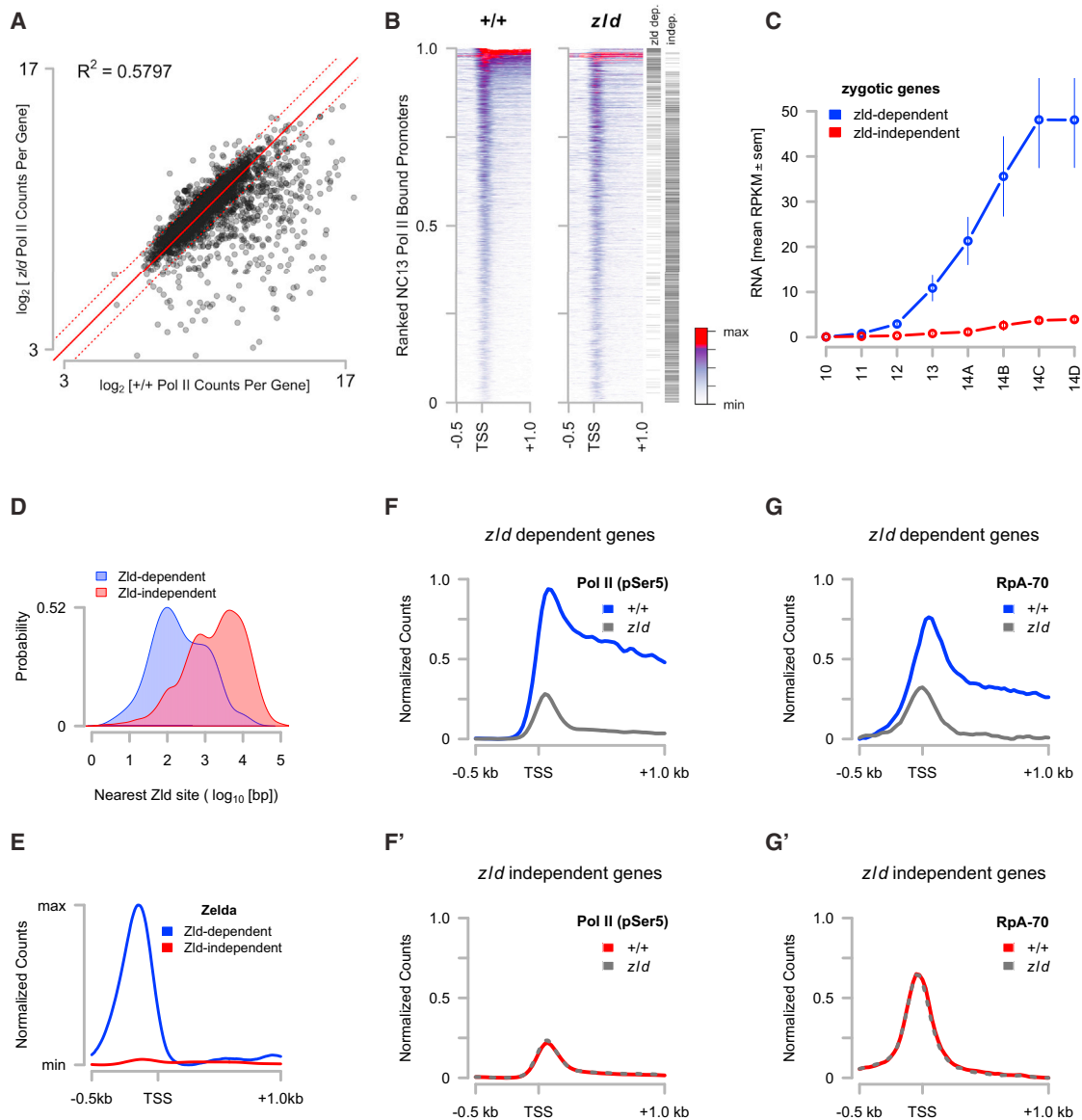
To test this prediction, we examined *mei-41* embryos heterozygously deficient for the transcription factor *Trithorax-like*/GAGA Factor known to be required for establishment of poised Pol II at heat-shock promoters and for embryonic transcriptional regulation (Bhat et al., 1996; Shopland et al., 1995). Embryos from *mei-41*; *Df(3L)ED4543/+* (*Df(Trl)*/+) mothers complete the syncytial divisions without mitotic catastrophe in 41% of cases following a short ( $13.2 \pm 0.36$  min) NC13 (Figures 7A and 7B), ultimately yielding hatching larvae (13%,  $n = 199$ ). Similar hatch rates are obtained using both an overlapping deficiency (*Df(3L) fz-M21*) and an allele of *Trl* (81.1, data not shown). Additionally, *Df(Trl)*/+ embryos otherwise wild-type for *mei-41* demonstrate a moderately shortened NC13 time ( $17.2 \pm 0.44$  min) compared with wild-type embryos (Figure 7A). Taken together, we conclude that the initial phases of ZGA trigger the MBT replication checkpoint, and the conflict between ZGA and DNA replication can be mitigated by reducing transcriptional initiation without a corresponding effect on cell-cycle duration.

## DISCUSSION

On the basis of five central observations, we conclude that the MBT replication checkpoint is activated in response to the de

novo recruitment of Pol II to chromatin at NC13. First, zygotic DNA differs in its capacity to trigger the checkpoint, correlating not with total DNA content, but rather with the quantity of transcriptionally engaged loci. Second, checkpoint activation coincides with large-scale de novo recruitment of Pol II throughout the genome. Third, sites of replication stalling as measured by Rpa-70 localize to transcribed regions of the genome. Fourth, reduced Pol II occupancy in *zelda* germline clones results in a local reduction of the genomic occupancy of Rpa-70 at *zelda* targets. Fifth, reducing total Pol II occupancy at NC13 suppresses the lethality associated with defects in the replication checkpoint. Our results therefore suggest a simple model for the coordination of zygotic genome activation and cell-cycle remodeling downstream of N:C ratio measurement.

Central to the concept of the MBT are the timing mechanisms that coordinate changes in maternal/zygotic RNA expression and cell-cycle behavior. In our model, the MBT replication checkpoint coordinates ZGA with cell-cycle remodeling, responding to large-scale transcriptional engagement to initiate changes in the maternal cell cycle. In this sense, the replication checkpoint is an indirect “sensor” of the N:C ratio, responding instead to a proxy of nuclear content in the form of the fraction of the zygotic genome engaged in transcription. We therefore propose that cell-cycle remodeling at the *Drosophila* MBT is zygotically driven by a two-step mechanism. First, the replication checkpoint, in response to de novo Pol II recruitment, drives Chk1-dependent downregulation of Cdc25 catalytic activity (Edgar and Datar, 1996; Peng et al., 1997), leading to attenuation of Cdk1 kinase activity and transient cell-cycle lengthening. Next, several zygotic genes drive the specific proteolytic degradation of the Cdc25 homolog Twine during NC14 (Di Talia et al., 2013; Farrell and O’Farrell, 2013). The resultant downregulation of Cdk1 activity leads to the acquisition of several hallmarks of the zygotic cell cycle, including a G2 phase and early and late



**Figure 6. Loss of Pol II Binding Reduces RPA-70 Binding to Transcribed Regions**

(A)  $\log_2$ [Pol II CPM] values for genes in the set of NC13-bound promoters were plotted for both wild-type and for *zld*<sup>294</sup> NC13 stage embryos. The solid red line indicates no change between samples and the dotted red lines indicate 2-fold changes in either direction.

(B) Promoter-proximal Pol II counts for both wild-type (+/+) and *zld*<sup>294</sup> were plotted as in Figure 2A. The position of each promoter in the classes of “zelda dependent” or “zelda independent” loci are marked by a black hashmark on the right margin.

(C) RPKM values for genes in the “zelda dependent” (blue) and the “zelda independent” (red) sets were extracted from Lott et al. (2011) and averaged. Mean RPKM values  $\pm$  SEM are plotted from NC10 through NC14.

(D) Kernel density estimates for distances between a known Zelda protein binding site (from Harrison et al., 2011) and TSSs in the “zelda dependent” (blue) and “zelda independent” (red) classes ( $p < 0.01$ , Wilcoxon rank sum test).

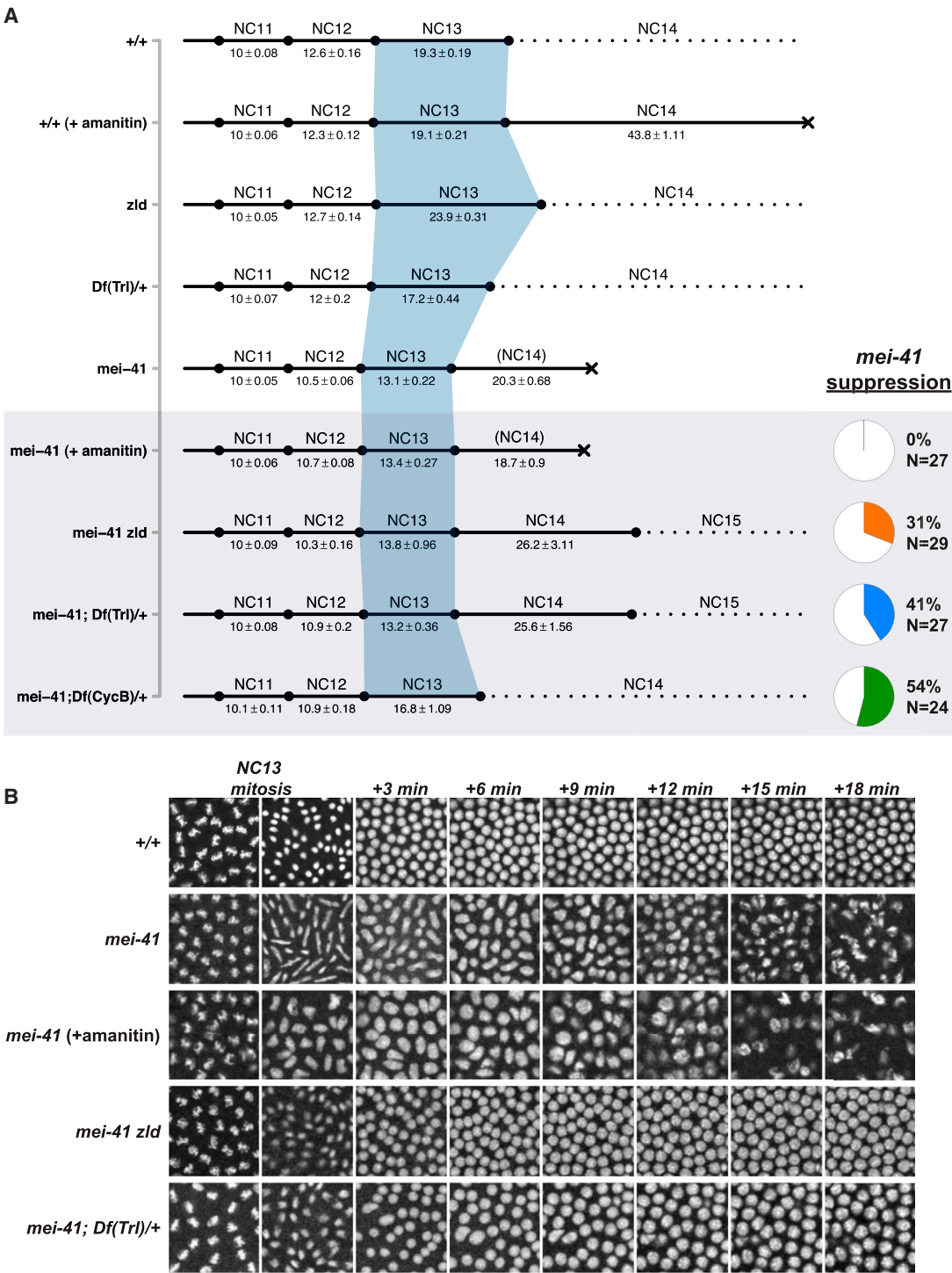
(E) The promoter proximal distribution of Zelda protein (from Harrison et al., 2011) for genes in the “zelda dependent” (blue) and “zelda independent” (red) sets was plotted.

(F and F') The NC13 promoter proximal distribution of Pol II for zelda-dependent loci (F) was plotted for wild-type (blue) and *zelda* (gray) embryos. (F') shows the distribution of Pol II at zelda independent loci for wild-type (red) and *zelda* (gray-dashed) embryos.

(G and G') The NC13 promoter proximal distribution of RPA-70 for zelda-dependent loci (G) was plotted for wild-type (blue) and *zelda* (gray) embryos. (G') shows the distribution of RPA-70 at zelda independent loci for wild-type (red) and *zelda* (gray-dashed) embryos.

See also Table S3.





**Figure 7. Reduced Pol II Recruitment Suppresses *mei-41* Lethality**

(A) Syncytial cell-cycle times for the indicated genotypes/treatments were measured by time-lapse confocal microscopy of H2Av-GFP. Time is represented in minutes ± SEM. Lethality is represented by a black X. The gray box highlights conditions tested for suppression of *mei-41* lethality. Data for suppressing individuals only are shown for the final three genotypes. The pie charts at bottom right indicate the frequency of *mei-41* suppression for the associated genotypes. Wild-type and *mei-41* data are reproduced from Figure 1. n = 27 (+/+ (+amanitin)), n = 30 (*zld* germline clones), n = 20 (*Df(Trl)/+*).

(B) Representative images (2,500 μm<sup>2</sup>) from time-lapse recordings from (A) are shown in 3 min intervals beginning at metaphase 13 through 18 min into NC14. Note the absence of defective NC13 mitosis in *mei-41 zld* and *mei-41; Df(Trl)/+* and subsequent wild-type nuclear morphology compared with *mei-41* alone or *mei-41* (+amanitin).

replicating chromatin domains (Farrell et al., 2012). In this model, cell-cycle remodeling is initiated by checkpoint-dependent regulation of catalytic levels but ultimately completed and stabilized by zygotic gene activity.

Our model therefore predicts that characterizing the control of Pol II recruitment to chromatin will elucidate how the N:C ratio timer ultimately drives cell-cycle remodeling. For at least a subset of zygotic genes, the onset of transcription correlates with the duration of interphase (Edgar and Schubiger, 1986). Interphase length is itself controlled by Cyclin/Cdk activity, and Cyclin dosage is gradually titrated by increasing nuclear content, resulting in a gradual checkpoint-independent lengthening of the syncytial cell cycle (Edgar et al., 1994; Ji et al., 2004). Therefore, N:C ratio-dependent ZGA could be activated once interphase time advances beyond a critical length. In addition, one or more uncharacterized N:C-ratio independent timers drive maternal mRNA clearance and activation of the class of time- or stage-dependent zygotic transcripts (Benoit et al., 2009; Lu et al., 2009; Tadros et al., 2003). Since expression of both N:C ratio-dependent and independent classes of zygotic transcripts is prevented by blocking translation before syncytial blastoderm stages (Edgar and Schubiger, 1986), one possible mechanism for timing events independently of the N:C-ratio is regulated translation of essential factors such as *smaug* and *zelda* (Benoit et al., 2009; Harrison et al., 2010). Indeed, the class of *zld*-dependent genes is enriched for the class of time/stage-dependent zygotic genes (Lu et al., 2009), supporting the idea that *zelda* drives N:C ratio independent ZGA. These observations support the emerging idea that ZGA is driven not by any one discrete mechanism, but rather by a collection of different, yet synchronized, systems.

One important question for future investigation will be to define the features of ZGA that trigger the MBT replication checkpoint. Our work suggests that the trigger of the checkpoint is upstream of entry into productive transcriptional elongation. Importantly, two mutants that confer premature ZGA have a corresponding premature activation of the MBT replication checkpoint (Pérez-Montero et al., 2013; Sung et al., 2013). A mutant of the early embryonic linker histone *Bigh1*, for example, causes early ZGA in the presence of widespread DNA damage (Pérez-Montero et al., 2013). Similarly, a mutation of the large subunit of RNA Polymerase II (*Rpl1215<sup>X161</sup>*) also causes premature ZGA, triggering an early replication checkpoint and cell-cycle pause (Sung et al., 2013). Although the precise nature of these mutant phenotypes is not yet clear, it is possible that they result from increased accessibility of Pol II to pre-MBT chromatin. These phenotypes are consistent with a model where the MBT replication checkpoint scales with zygotic transcriptional engagement.

It is also important to note that we have not determined whether the form of the trigger is actual replication stress, or rather stress-independent recruitment of the Rpa complex to promoters. It is possible that Pol II occupancy represents a previously unseen “roadblock” to the DNA replisome (Azvolinsky et al., 2009), which can lead to replication stress (reviewed in Bermejo et al., 2012). However, evidence from other model systems supports a stress-independent pathway. In budding yeast, Rpa binds to promoters and actively transcribed genes independently of the DNA replisome (Sikorski et al., 2011). We show that Rpa-70 binds to both active and poised promoters at NC13, and further

evidence suggests that Rpa could be recruited as part of the Pol II complex itself (Maldonado et al., 1996) or even function as an essential component of poised chromatin architecture (Fujimoto et al., 2012). In the latter example, interaction with the Rpa complex is necessary for HSF1 binding and for pre-loading of RNA Pol II at heat-shock promoters by recruiting the histone chaperone FACT (Fujimoto et al., 2012). Although proteomic screens have not identified an HSF-Rpa interaction in *Drosophila*, Rpa does appear to interact physically with GAGA-binding proteins Pipsqueak and Trithorax-like (Guruharsha et al., 2011), the latter of which we show to interact genetically with *mei-41* (Figure 7). Therefore, it remains possible that the mechanism driving engagement of Pol II itself involves large-scale recruitment of the Rpa complex to chromatin, thus mimicking a signal of replication stress to activate the checkpoint.

## EXPERIMENTAL PROCEDURES

Complete experimental procedures are included in the [Supplemental Information](#).

### Measurement of NC13 Duration

Three different crosses were used to generate embryos with 76%–124% DNA content: [C(1)RM/0; Rpl135 EGFP/+ x C(1;Y)1/0], [w; Rpl135 EGFP/+ x C(1;Y)1/0], and [C(1)DX/Y; Rpl135 EGFP/+ x C(1;Y)1/0]. Genotypes were scored by counting the number of Rpl135 EGFP foci at NC13 (e.g., as in Figure 1B) and by scoring NC14 nullo-X phenotypes of 0/0 and Y/0 embryos. Wild-type male and female embryos were distinguished by scoring zygotic expression of a paternally supplied X-linked GFP transgene (= X/X).

Up to 15 embryos were dechorionated and affixed to a glass coverslip, overlaid with halocarbon oil, and simultaneously imaged by laser scanning confocal microscopy at a 30-s frame rate per embryo. Cell-cycle times were scored as the duration between successive anaphases. To control for day-to-day fluctuations in room temperature, cell-cycle times were normalized by setting the mean NC11 time to 10 min and scaling NC12 and 13 accordingly based on the mean NC11 time of all embryos on the slide. The NC13 times reported in Figure 1 are based on time-lapse recordings from 182 embryos. (n = 11 [X/Y, Y/0 and 0/0], n = 12 [X/X], n = 13 [X/Y/0], n = 15 [XX/X], n = 17 [Y/Y], n = 18 [XX/0], n = 20 [XX<sup>bb</sup>/0 and X/0], and n = 21 [XX/Y and XX<sup>bb</sup>/Y]).

### Scoring *mei-41* Suppression

Embryos from *mei-41<sup>29D</sup> f zld<sup>294</sup> FRT19A; H2Av-GFP/+* germline clones were used to score *zld* suppression of *mei-41*. *mei-41<sup>D3/29D</sup>*; Df(3L)ED4543/H2Av-GFP was the maternal genotype for scoring *Trf* suppression and *mei-41<sup>D3/29D</sup>*; Df(2R)59AB/+; H2Av-GFP/+ was the maternal genotype for scoring *CycB* suppression. Embryos were imaged as described above. The *mei-41* phenotype was scored as “suppressed” if >75% of imaged nuclei successfully completed mitosis 13 (and 14 in cases of extra divisions) without evidence of anaphase bridging and if wild-type blastoderm morphology was maintained during post-MBT cell-cycle pause.

### Chromatin Immunoprecipitation

Embryos were crosslinked for 15 min in a solution of 2 ml PBS + 0.5% Triton X-100 overlaid with 6 ml Heptane and 180  $\mu$ l 20% fresh paraformaldehyde. Interphase embryos of specific stages were sorted under an epifluorescent dissection microscope on the basis of nuclear density by means of the Rpa-70 EGFP transgene. Subsets of NC14 embryos were collected by measurement of nuclear elongation on a compound microscope with 20 $\times$  objective. ChIP was performed essentially as described in (Blythe et al., 2009), with modifications noted in the [Supplemental Information](#).

### Sequencing and Analysis

Single-end sequencing of barcoded libraries was performed by the Lewis Sigler Institute for Integrative Genomics Sequencing Core Facility on an

Illumina HiSeq 2500 with read length of 67 bp. Libraries were prepared with the NEBNext ChIP-seq library prep mastermix kit (NEB) according to the manufacturer's instructions. All data reflect the mean of two independent biological replicates.

Sequences were mapped to the *Drosophila* genome (dm3) using default settings on Bowtie (Langmead et al., 2009). Regions of significant enrichment were determined using Zinba (Rashid et al., 2011), differential binding was determined using edgeR (Robinson et al., 2010), and all other analysis was performed using the GenomicRanges package in R (Lawrence et al., 2013) (<http://www.R-project.org/>). Sequences and peaks mapping to chrU and Uextra were not considered. Regions of enrichment were mapped to a modified Ensembl transcript database by identifying peaks within 1 kb of an annotated TSS, excluding transcripts <125 bp in length. The mean CPM values for 25 bp windows across the length of the genome were calculated and used to determine additional comparisons described in the text.

### ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the ChIP-seq data reported in this paper is GSE62925.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, one figure, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.050>.

### AUTHOR CONTRIBUTIONS

S.A.B. designed and performed all experiments and analysis. E.F.W. discussed experimental results and performed difficult crosses. S.A.B. and E.F.W. wrote the manuscript.

### ACKNOWLEDGMENTS

We thank the Bloomington *Drosophila* Stock Center for fly stocks. We thank C. Hannon, A. Roknabadi, and B. Pelham-Webb for help with experiments, all members of the E.F.W. and Schubach laboratories, especially S. Little and B. He for lively discussion. We thank R. Kadzik, P. Klein, T. Schubach, J. Lieb, M. Harrison, J. Zeitlinger, T. Orr-Weaver, W. Sullivan, and P. O'Farrell for helpful comments and C. Rushlow, P. Schedl, and J. Sekelsky for fly stocks. We are indebted to the staff of the Sequencing Core Facility of the Lewis Sigler Institute and to L. Parsons for support for the ChIP-seq studies. We thank S. Di Talia for comments on the manuscript. This work was supported in part by grant 5R37HD15587 from the National Institute of Child Health and Human Development (NICHD) to E.F.W. and Ruth Kirschstein National Research Service Award (NRSA) Postdoctoral Fellowship 1F32HD072653 from NICHD to S.A.B. E.F.W. is an investigator with the Howard Hughes Medical Institute.

Received: July 15, 2014

Revised: November 3, 2014

Accepted: January 8, 2015

Published: March 5, 2015

### REFERENCES

Azvolinsky, A., Giresi, P.G., Lieb, J.D., and Zakian, V.A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in *Saccharomyces cerevisiae*. *Mol. Cell* 34, 722–734.

Benoit, B., He, C.H., Zhang, F., Votruba, S.M., Tadros, W., Westwood, J.T., Smibert, C.A., Lipshitz, H.D., and Theurkauf, W.E. (2009). An essential role for the RNA-binding protein Smaug during the *Drosophila* maternal-to-zygotic transition. *Development* 136, 923–932.

Bermejo, R., Lai, M.S., and Foiani, M. (2012). Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Mol. Cell* 45, 710–718.

Bhat, K.M., Farkas, G., Karch, F., Gyurkovics, H., Gausz, J., and Schedl, P. (1996). The GAGA factor is required in the early *Drosophila* embryo not only for transcriptional regulation but also for nuclear division. *Development* 122, 1113–1124.

Blythe, S.A., Reid, C.D., Kessler, D.S., and Klein, P.S. (2009). Chromatin immunoprecipitation in early *Xenopus laevis* embryos. *Dev. Dyn.* 238, 1422–1432.

Brodsky, M.H., Sekelsky, J.J., Tsang, G., Hawley, R.S., and Rubin, G.M. (2000). *mus304* encodes a novel DNA damage checkpoint protein required during *Drosophila* development. *Genes Dev.* 14, 666–678.

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., and Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *eLife* 2, e00861.

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Titration of four replication factors is essential for the *Xenopus laevis* midblastula transition. *Science* 341, 893–896.

Conn, C.W., Lewellyn, A.L., and Maller, J.L. (2004). The DNA damage checkpoint in embryonic cell cycles is dependent on the DNA-to-cytoplasmic ratio. *Dev. Cell* 7, 275–281.

Crest, J., Oxnard, N., Ji, J.-Y., and Schubiger, G. (2007). Onset of the DNA replication checkpoint in the early *Drosophila* embryo. *Genetics* 175, 567–584.

Dasso, M., and Newport, J.W. (1990). Completion of DNA replication is monitored by a feedback system that controls the initiation of mitosis in vitro: studies in *Xenopus*. *Cell* 61, 811–823.

Di Talia, S., She, R., Blythe, S.A., Lu, X., Zhang, Q.F., and Wieschaus, E.F. (2013). Posttranslational control of Cdc25 degradation terminates *Drosophila*'s early cell-cycle program. *Curr. Biol.* 23, 127–132.

Edgar, B.A., and Datar, S.A. (1996). Zygotic degradation of two maternal Cdc25 mRNAs terminates *Drosophila*'s early cell cycle program. *Genes Dev.* 10, 1966–1977.

Edgar, B.A., and Schubiger, G. (1986). Parameters controlling transcriptional activation during early *Drosophila* development. *Cell* 44, 871–877.

Edgar, B.A., Sprenger, F., Duronio, R.J., Leopold, P., and O'Farrell, P.H. (1994). Distinct molecular mechanism regulate cell cycle timing at successive stages of *Drosophila* embryogenesis. *Genes Dev.* 8, 440–452.

Farrell, J.A., and O'Farrell, P.H. (2013). Mechanism and regulation of Cdc25/Twine protein destruction in embryonic cell-cycle remodeling. *Curr. Biol.* 23, 118–126.

Farrell, J.A., Shermoen, A.W., Yuan, K., and O'Farrell, P.H. (2012). Embryonic onset of late replication requires Cdc25 down-regulation. *Genes Dev.* 26, 714–725.

Fasulo, B., Koyama, C., Yu, K.R., Homola, E.M., Hsieh, T.S., Campbell, S.D., and Sullivan, W. (2012). Chk1 and Wee1 kinases coordinate DNA replication, chromosome condensation, and anaphase entry. *Mol. Biol. Cell* 23, 1047–1057.

Foe, V.E., and Alberts, B.M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* 61, 31–70.

Fogarty, P., Kalpin, R.F., and Sullivan, W. (1994). The *Drosophila* maternal-effect mutation grapes causes a metaphase arrest at nuclear cycle 13. *Development* 120, 2131–2142.

Fujimoto, M., Takaki, E., Takii, R., Tan, K., Prakasam, R., Hayashida, N., Iemura, S., Natsume, T., and Nakai, A. (2012). RPA assists HSF1 access to nucleosomal DNA by recruiting histone chaperone FACT. *Mol. Cell* 48, 182–194.

Gong, X.Q., Nedialkov, Y.A., and Burton, Z.F. (2004). Alpha-amanitin blocks translocation by human RNA polymerase II. *J. Biol. Chem.* 279, 27422–27427.

Gururharsha, K.G., Rual, J.-F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O., et al. (2011). A protein complex network of *Drosophila melanogaster*. *Cell* 147, 690–703.

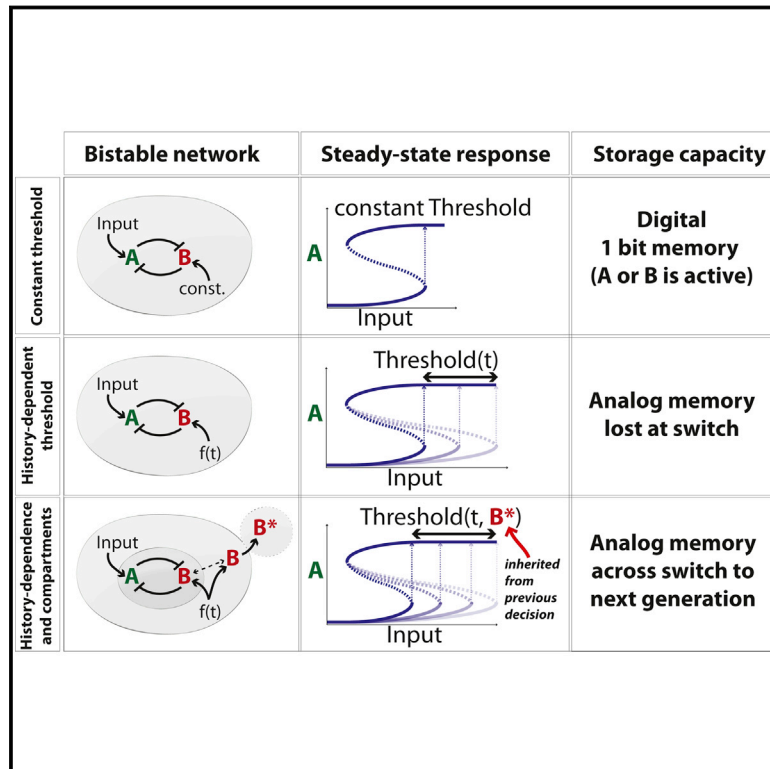
Harrison, M.M., Botchan, M.R., and Cline, T.W. (2010). Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed *Drosophila* genes. *Dev. Biol.* 345, 248–255.

- Harrison, M.M., Li, X.Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet.* 7, e1002266.
- Ji, J.-Y., Squirrell, J.M., and Schubiger, G. (2004). Both cyclin B levels and DNA-replication checkpoint control the early embryonic mitoses in *Drosophila*. *Development* 131, 401–411.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- LaRocque, J.R., Dougherty, D.L., Hussain, S.K., and Sekelsky, J. (2007). Reducing DNA polymerase alpha in the absence of *Drosophila* ATR leads to P53-dependent apoptosis and developmental defects. *Genetics* 176, 1441–1451.
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.
- Li, B., Weber, J.A., Chen, Y., Greenleaf, A.L., and Gilmour, D.S. (1996). Analyses of promoter-proximal pausing by RNA polymerase II on the hsp70 heat shock gene promoter in a *Drosophila* nuclear extract. *Mol. Cell. Biol.* 16, 5433–5443.
- Li, X.Y., Harrison, M.M., Villalta, J.E., Kaplan, T., and Eisen, M.B. (2014). Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *eLife* 3, 3.
- Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* 456, 400–403.
- Long, E.O., and Dawid, I.B. (1980). Repeated genes in eukaryotes. *Annu. Rev. Biochem.* 49, 727–764.
- Lott, S.E., Villalta, J.E., Schroth, G.P., Luo, S., Tonkin, L.A., and Eisen, M.B. (2011). Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. *PLoS Biol.* 9, e1000590.
- Lu, X., Li, J.M., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2009). Coupling of zygotic transcription to mitotic control at the *Drosophila* mid-blastula transition. *Development* 136, 2101–2110.
- Maldonado, E., Shiekhattar, R., Sheldon, M., Cho, H., Drapkin, R., Rickert, P., Lees, E., Anderson, C.W., Linn, S., and Reinberg, D. (1996). A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature* 381, 86–89.
- Nien, C.-Y., Liang, H.-L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J.R., and Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet.* 7, e1002339.
- O'Farrell, P.H., Stumpff, J., and Su, T.T. (2004). Embryonic cleavage cycles: how is a mouse like a fly? *Curr. Biol.* 14, R35–R45.
- Peng, C.Y., Graves, P.R., Thoma, R.S., Wu, Z., Shaw, A.S., and Piwnicka-Worms, H. (1997). Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. *Science* 277, 1501–1505.
- Pérez-Montero, S., Carbonell, A., Morán, T., Vaquero, A., and Azorín, F. (2013). The embryonic linker histone H1 variant of *Drosophila*, dBigH1, regulates zygotic genome activation. *Dev. Cell* 26, 578–590.
- Rashid, N.U., Giresi, P.G., Ibrahim, J.G., Sun, W., and Lieb, J.D. (2011). ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions. *Genome Biol.* 12, R67.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- Shermoen, A.W., McClelland, M.L., and O'Farrell, P.H. (2010). Developmental control of late replication and S phase length. *Curr. Biol.* 20, 2067–2077.
- Shopland, L.S., Hirayoshi, K., Fernandes, M., and Lis, J.T. (1995). HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. *Genes Dev.* 9, 2756–2769.
- Sibon, O.C., Stevenson, V.A., and Theurkauf, W.E. (1997). DNA-replication checkpoint control at the *Drosophila* midblastula transition. *Nature* 388, 93–97.
- Sibon, O.C., Laurençon, A., Hawley, R., and Theurkauf, W.E. (1999). The *Drosophila* ATM homologue Mei-41 has an essential checkpoint function at the midblastula transition. *Curr. Biol.* 9, 302–312.
- Sikorski, T.W., Ficarro, S.B., Holik, J., Kim, T., Rando, O.J., Marto, J.A., and Buratowski, S. (2011). Sub1 and RPA associate with RNA polymerase II at different stages of transcription. *Mol. Cell* 44, 397–409.
- Sung, H.W., Spangenberg, S., Vogt, N., and Großhans, J. (2013). Number of nuclear divisions in the *Drosophila* blastoderm controlled by onset of zygotic transcription. *Curr. Biol.* 23, 133–138.
- Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. *Development* 136, 3033–3042.
- Tadros, W., Houston, S.A., Bashirullah, A., Cooperstock, R.L., Semotok, J.L., Reed, B.H., and Lipshitz, H.D. (2003). Regulation of maternal transcript destabilization during egg activation in *Drosophila*. *Genetics* 164, 989–1001.
- Wold, M.S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* 66, 61–92.
- Zeitlinger, J., Stark, A., Kellis, M., Hong, J.-W., Nechaev, S., Adelman, K., Levine, M., and Young, R.A. (2007). RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat. Genet.* 39, 1512–1516.
- Zou, L., and Elledge, S.J. (2003). Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300, 1542–1548.



# Compartmentalization of a Bistable Switch Enables Memory to Cross a Feedback-Driven Transition

## Graphical Abstract



## Authors

Andreas Doncic, Oguzhan Atay, ...,  
Mart Loog, Jan M. Skotheim

## Correspondence

skotheim@stanford.edu

## In Brief

The spatial organization of the G1/S switch enables the intergenerational transmission of memory of pheromone exposure in budding yeast.

## Highlights

- Yeast decide to enter the cell cycle based on the history of pheromone exposure
- Compartmentalization enables transmission of memory from mother to daughter
- Intergenerational memory of pheromone exposure is stored as cytoplasmic Far1
- Anchoring of cytoplasmic Far1 by Cdc24 is required for intergenerational memory



# Compartmentalization of a Bistable Switch Enables Memory to Cross a Feedback-Driven Transition

Andreas Doncic,<sup>1</sup> Oguzhan Atay,<sup>1</sup> Ervin Valk,<sup>2</sup> Alicia Grande,<sup>3</sup> Alan Bush,<sup>3</sup> Gustavo Vasen,<sup>3</sup> Alejandro Colman-Lerner,<sup>3</sup> Mart Loog,<sup>2</sup> and Jan M. Skotheim<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Stanford University, Stanford, CA 94305, USA

<sup>2</sup>Institute of Technology, University of Tartu, 50411, Estonia

<sup>3</sup>IFIBYNE-UBA-CONICET and Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires C1428EHA, Argentina

\*Correspondence: [skotheim@stanford.edu](mailto:skotheim@stanford.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.032>

## SUMMARY

Cells make accurate decisions in the face of molecular noise and environmental fluctuations by relying not only on present pathway activity, but also on their memory of past signaling dynamics. Once a decision is made, cellular transitions are often rapid and switch-like due to positive feedback loops in the regulatory network. While positive feedback loops are good at promoting switch-like transitions, they are not expected to retain information to inform subsequent decisions. However, this expectation is based on our current understanding of network motifs that accounts for temporal, but not spatial, dynamics. Here, we show how spatial organization of the feedback-driven yeast G1/S switch enables the transmission of memory of past pheromone exposure across this transition. We expect this to be one of many examples where the exquisite spatial organization of the eukaryotic cell enables previously well-characterized network motifs to perform new and unexpected signal processing functions.

## INTRODUCTION

Cellular signaling pathways are used to transmit information about the extra- and intra-cellular environment. Specific outputs from such signaling pathways are then used by decision-making networks to determine cellular response. Currently, signaling pathways are most often described as static schematics based on a combination of genetic dependencies and biochemical interactions. While a good first step, such a characterization can neither describe nor predict the pathway dynamics that determine cellular response to time-dependent input signals (Behar et al., 2008; Yosef and Regev, 2011). Indeed, outputs of the regulatory networks controlling proliferation and apoptosis depend on the history of dynamic input signals, not only on current levels (Doncic and Skotheim, 2013; Lee et al., 2012; Purvis et al., 2012). This strongly suggests that the ability to retain information from prior states is a key determinant informing cellular decision making.

Signaling dynamics play important roles in many networks regulating switch-like transitions between distinct states. The switch-like nature of transitions often arises from positive feedback loops that quickly increase the activity of key regulatory proteins when triggered by input signals above a specific threshold. Networks containing positive feedback loops frequently give rise to bistability, i.e., for a range of input signals, the output will be one of two possible values depending on the history of the input signal. However, this is a very simple form of history dependence as all possible time-dependent input signals get mapped onto only two possible outputs, i.e., history dependence is collapsed onto only a single bit of information. This implies that while positive feedback loops may be good at promoting switch-like transitions, they appear unable to retain more than rudimentary information about signaling pathway history. It is therefore improbable that a positive-feedback-driven switch can be used to transmit information to inform future cellular decisions. However, this conclusion is based on the current framework for analyzing network motifs such as feedback-loops or feed-forward interactions (Alon, 2007), which accounts for temporal but not spatial dynamics. Thus, while it is well-known that spatial organization plays an important role in signal transduction, we do not currently know how or if the eukaryotic cell's spatial organization can affect existing motif functions or give rise to entirely new motif functions (Howell et al., 2012; Kholodenko et al., 2010; Santos et al., 2012).

To better understand how spatial organization might affect cellular signal processing, we decided to examine the cell-cycle control network responsible for the decision to divide in budding yeast. In yeast, the decision to commit to cell division takes place in late G1, prior to DNA replication at a point called *Start* (Hartwell et al., 1974). Multiple internal and external signals are integrated to determine when a cell passes *Start*, beyond which cells no longer respond to mating pheromone ( $\alpha$ -factor). *Start* is a switch-like, irreversible transition that corresponds to the activation of a positive feedback loop of cyclin-dependent kinase (Cdk1) activity (Doncic et al., 2011). Specifically, Cln3-Cdk partially inactivates Whi5, a transcriptional inhibitor of the expression of the G1 cyclins *CLN1* and *CLN2*. The expression of Cln1 and Cln2 complete inactivation of Whi5 by forming a positive feedback loop (Costanzo et al., 2004; de Bruin et al., 2004; Skotheim et al., 2008).

Prior to *Start*, cells can be arrested by pheromone-dependent activation of the mitogen activated protein kinase (MAPK) mating pathway (Chen and Thorner, 2007). Upon pheromone exposure, the MAPK Fus3 phosphorylates and activates the Cdk inhibitor Far1, which inhibits the G1 cyclins essential for progression through *Start* (Chang and Herskowitz, 1990; Gartner et al., 1998; Jeoung et al., 1998; Peter et al., 1993; Pope et al., 2014; Tyers and Futcher, 1993). Conversely, post-*Start*, the G1 cyclins inhibit the mating pathway by targeting the upstream scaffold protein Ste5 as well as Far1 (Garrenton et al., 2009; Henchoz et al., 1997; Peter and Herskowitz, 1994; Strickfaden et al., 2007; Tyers and Futcher, 1993) (Figure S1A). Thus, progression through *Start* drives an increase in cyclin expression that results in Far1 degradation, whereas pre-*Start* exposure to pheromone leads to Far1 activation, G1 cyclin inhibition, and G1 arrest (Doncic et al., 2011; McKinney et al., 1993; Pope et al., 2014). In other words, the regulatory network underlying *Start* is bistable, where a well-defined commitment point separates stable low- and high-Cdk activity states, and only the low-Cdk activity state can be inhibited by MAPK signaling (Doncic et al., 2011).

Although this characterization of *Start* is accurate for a step input of high pheromone concentration, cells exposed to low or intermediate pheromone concentrations do not arrest permanently, but rather delay progression through G1 (Hao et al., 2008; Malleshaiah et al., 2010; Moore, 1984). This suggests a more complex decision making machinery that balances the benefits of successful mating with the costs of staying arrested and both failing to mate and proliferate. Thus, while the *Start* network remains bistable, its output changes from a digital response to arrest or not, to an analog computation determining how long to arrest before reentering the cell division cycle. We previously showed that in this analog computation, yeast cells decide to reenter the cell cycle based on their history of exposure to pheromone during an arrest, not just the current pathway activity (Doncic and Skotheim, 2013). Time-dependent pheromone signals are processed by the MAPK pathway using a coherent feed-forward motif in which the MAPK Fus3 activates Far1 both by direct phosphorylation and by increasing its expression via the Ste12 transcription factor (Chang and Herskowitz, 1990; Errede and Ammerer, 1989; Gartner et al., 1998) (Figure S1A; red arrows). This architecture allows a robust yet rapidly reversible cellular state. Far1 accumulates to provide a memory so that cells exposed to pheromone for longer durations have more Far1 rendering them more reluctant to reenter the cell cycle. In addition, fast dephosphorylation allow Far1 to be rapidly inactivated so that cells can rapidly reenter the cell cycle if the MAPK signal plummets (Doncic and Skotheim, 2013).

Although the accumulation of Far1 provides a mechanism to remember the history of pheromone exposure during a single arrest, it does not suggest a mechanism to transmit this information to subsequent generations after cell-cycle reentry. This is because the mutual inhibition of Cdk and Far1 activity underlying the bistable *Start* switch is expected to target all Far1 for degradation once the cell cycle has been reentered. Similarly, the sharp switch at mitotic exit also employs ultra-sensitive protein degradation (Yang and Ferrell, 2013). Protein degradation may be useful to sharpen switches and reset regulatory circuits, but

comes at the cost of losing cellular memory. Thus, while bistable regulatory networks are excellent at generating all-or-none transitions, they limit the amount of information that can be propagated across these transitions.

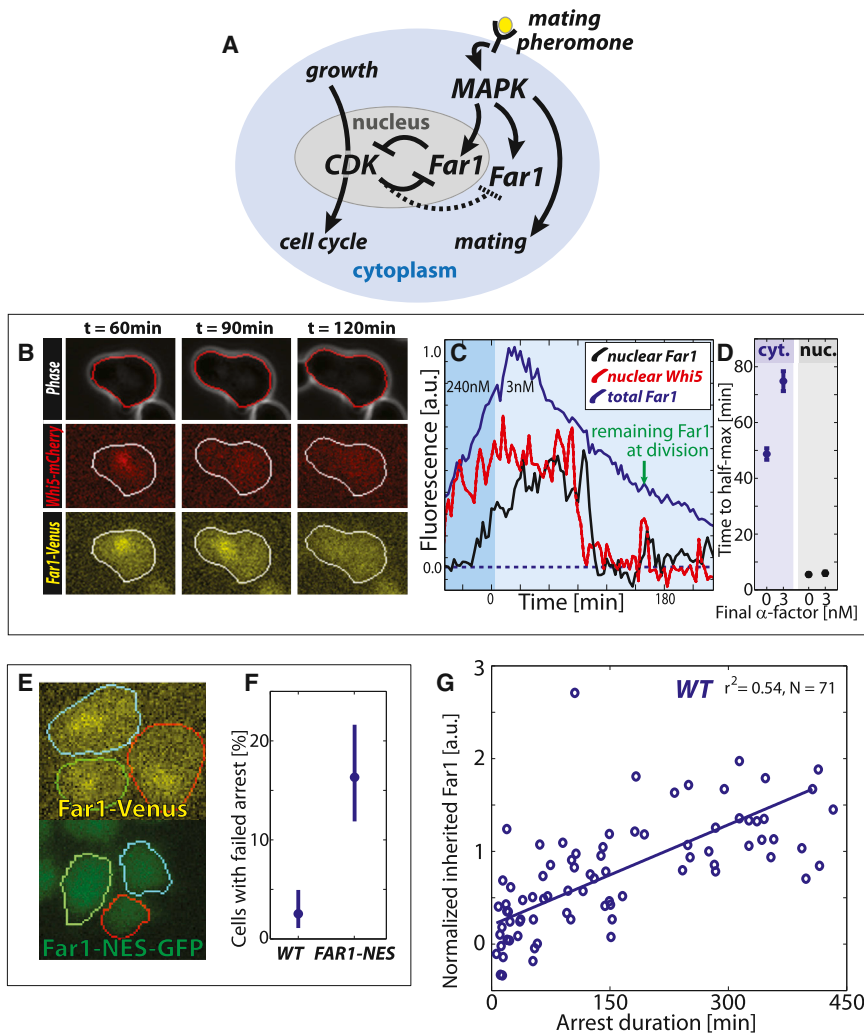
Here, we show how compartmentalization of the bistable G1 control network allows cellular memory to traverse the *Start* switch. Far1 is split into nuclear and cytoplasmic pools that combat distinct sets of cyclin-Cdk complexes allowing these two compartments of the Far1-Cdk switch to have distinct dynamics. Upon reentering the cell cycle from pheromone arrest, nuclear Far1 is rapidly degraded, while cytoplasmic Far1 is degraded much more slowly so that a substantial pool remains at the beginning of the next division cycle. We show that this inherited pool contributes to cell-cycle arrest in the daughter cells so that the mother cells are able to transmit their memory of pheromone exposure to the next generation. This intergenerational memory depends on the anchoring of Far1 to cytoplasmic Cdc24, a regulator of cell polarization. Thus, we demonstrate how compartmentalization of a bistable regulatory circuit enables an entirely new function to be performed by this well-characterized signaling motif. More broadly, our results argue that spatial organization can greatly enhance the function of regulatory motifs and is therefore just as integral to pathway function as network topology and chemical kinetics.

## RESULTS

### Nuclear Far1 and Nuclear Cln2 Function in Cell-Cycle Commitment

To determine if and how signal information could be propagated across a bistable switch, we examined the network regulating *Start*, the point of commitment to cell division in budding yeast (Figure 1A). Since cyclin-Cdk complexes phosphorylate Far1 to target it for degradation, we expected that Far1 would be rapidly degraded upon progression through *Start*.

To examine the localization and dynamics of Far1, we used a Far1-Venus fusion protein expressed from the endogenous locus (Figure 1B). This *FAR1-Venus* strain exhibited the same arrest kinetics as an unlabeled WT strain, and we will subsequently refer to *FAR1-Venus* strains as WT (Doncic and Skotheim, 2013). Unless specified otherwise, all strains are in a background lacking the Bar1 protease that cleaves mating pheromone (for strain and plasmid lists see Table S1 and Table S2). Cells were arrested in high pheromone (240 nM  $\alpha$ -factor) and released into pheromone-free medium using a previously described microfluidics-based assay (Doncic et al., 2011). Consistent with previous results (McKinney et al., 1993), Far1 was synthesized during mating arrest and mostly degraded post-*Start* after release into pheromone-free medium. However, the examination of Far1-Venus using time-lapse microscopy revealed a striking spatial dichotomy in Far1 degradation kinetics. The nuclear pool of Far1 is rapidly degraded in less than 10 min (approximately 7 min after *Start*), which is defined as when 50% of Whi5 has been exported from the nucleus (Doncic et al., 2011). Nuclear Far1 is degraded at approximately the same time as the Cdk-B-type cyclin inhibitor Sic1, which we previously measured as occurring ~8 min after *Start* (see Figures S1B and S1C for Far1 degradation



**Figure 1. Cytoplasmic Far1 Is Inherited to Provide Intergenerational Memory across the Start Switch**

(A) Schematic of the double-negative feedback (equivalent to positive feedback) network that regulates the switch between cell-cycle progression and pheromone arrest.

(B) Example images of segmented phase, Whi5-mCherry (red) and Far1-Venus (yellow) channels for cells reentering the cell cycle. Whi5-mCherry is nuclear in arrested cells.

(C) Example time series of nuclear Whi5-mCherry, and nuclear and cytoplasmic Far1-Venus that corresponds to the cell shown in (B). Nuclear Far1 is much more rapidly degraded than cytoplasmic Far1.

(D) Time from peak to half-maximum for cytoplasmic and nuclear Far1 in cells arrested in 240 nM and released into either 3 or 0 nM pheromone.

(E) Example *FAR1-Venus* and *FAR1-NES-GFP* cells arrested in 240 nM for 2 hr show that the nuclear localization is diminished in the *FAR1-NES-GFP* cells.

(F) A larger fraction of pre-Start *FAR1-NES* cells fails to arrest when abruptly exposed to 240 nM pheromone, where Start is defined as removal of 50% of nuclear Whi5-mCherry.

(G) Inherited Far1 in daughter cells is correlated with arrest duration.

Error bars in (D) denote SEM, while error bars in (F) denote 95% confidence intervals from 10,000 bootstrap iterations.

timing and (Doncic et al., 2011) for Sic1 degradation timing). This implies that Far1 degradation is likely coincident with the appearance of B-type cyclin activity in the nucleus. However, the cytoplasmic pool lingered and reached half-maximum  $\sim 50$  min after Start (Figures 1C and 1D). This observed difference in Far1 degradation kinetics may be due to the nuclear F-box protein Cdc4 that mediates Far1 degradation (Blondel et al., 2000). This demonstrates that there are two separate pools of Far1 protein being degraded on very different time scales.

The rapid degradation of nuclear Far1 upon progression through Start suggests that it is primarily this nuclear pool that contributes to the commitment decision in response to exposure to pheromone (Blondel et al., 1999; Blondel et al., 2000). To test this, we added a nuclear export sequence to the endogenous *FAR1* allele (*FAR1-NES*), which greatly reduced the nuclear pool without affecting expression levels (Figures 1E and S1D). We then examined the cellular response to an abrupt increase in pheromone in the framework we previously developed to examine Start (Doncic et al., 2011). When exposed to a step-increase of pheromone, pre-Start *FAR1-NES* cells were over six

times more likely than WT cells to fail to arrest despite not having traversed the Whi5-threshold (16% *FAR1-NES* versus 2.5% WT, Figure 1F). In addition, we found that nuclear, but not cytoplasmic Cln2 participated in Start (Figures S1E–S1I). Taken together, these results support a role for nuclear Far1 in Start.

### Cytoplasmic Far1 Provides Intergenerational Memory of Pheromone Exposure

Even though nuclear Far1 was important for Start, most Far1 in arrested cells ( $\sim 90\%$ ) is cytoplasmic and is not degraded rapidly upon cell-cycle reentry (Figures S2A–S2E). In fact, cytoplasmic Far1 is so slowly degraded after cell-cycle reentry that appreciable quantities are passed on to subsequent generations (Figures 1C and S2F). This is surprising because once cells reenter the cell cycle, these mother cells are desensitized to that level of pheromone and divide repeatedly without delay (Figure S2G) (Caudron and Barral, 2013; Doncic and Skotheim, 2013; Moore, 1984). To examine the role of inherited Far1 in daughter cells, we briefly arrested cells at high pheromone concentration (240 nM) before releasing the cells into an intermediate pheromone concentration (3 nM). The time to reach half-maximum Far1 post-Start in 3 nM pheromone was  $\sim 5$  min in the nuclear pool and  $\sim 75$  min in the cytoplasmic pool (Figure 1D). Thus, the time to reach cytoplasmic half-maximum post-Start is



increased relative to cells reentering in pheromone-free medium. Consequently, daughter cells entering the cell cycle in 3 nM pheromone inherited an increased amount of Far1 compared to daughter cells entering the cell cycle in pheromone-free medium (Figure S2H). This finding, that cells cycling in higher pheromone concentrations pass increasing amounts of Far1 on to their daughter cells, led us to hypothesize that inheritance of cytoplasmic Far1 is the molecular basis of an *intergenerational memory* of pheromone exposure.

To test the *intergenerational memory* hypothesis, we measured both the amount of inherited Far1 and subsequent G1 duration for daughter cells cycling in 3 nM, an intermediate mating pheromone concentration (see [Experimental Procedure](#) and Figure S2I). The more Far1 a daughter cell inherited, the longer it delayed progression through G1, supporting the hypothesis that mother cells transmit information about pheromone exposure to their daughters through cytoplasmic Far1 (Figure 1G). We also examined if differential inheritance of MAPK pathway scaffold Ste5, which affects pheromone signaling in a dosage-dependent manner (Thomson et al., 2011), could affect arrest duration, but found no effect (Figure S2J and S2K).

### Compartmentalization Is Supported by a Fixed Fraction of Cytoplasmic Far1

The rapid degradation of the nuclear, but not cytoplasmic pool of Far1, requires a slow exchange between these two pools. Indeed, it would be impossible to maintain Far1 post-*Start* if Far1 were exchanged rapidly between the two pools as the half-life of nuclear Far1 is ~5 min. To investigate this requirement, we photobleached the nucleus of *FAR1-Venus* cells and measured the recovery of nuclear fluorescence (Figure 2A). Pre-*Start* cells were identified by examining the localization of Whi5-mKO fusion proteins expressed from the endogenous locus. After photobleaching, a significant recovery of the nuclear fraction on the 10 s timescale was seen. However, the nuclear to cytoplasmic fluorescence ratio did not recover to its initial level and reached a plateau prior to 30 s (Figures 2B–2D). Note that there is little new protein synthesis or degradation over the time frame of the experiment so that nearly all the recovery is due to protein translocation. The incomplete recovery of the nuclear-to-cytoplasmic ratio of Far1-*Venus* indicates that there is a pool of Far1 molecules that does not shuttle between the nucleus and the cytoplasm. As a control, we examined recovery of the yellow fluorescent protein YFP expressed from an *ACT1* promoter. The YFP nuclear-to-cytoplasmic ratio completely recovered after bleaching, which is consistent with rapid and unencumbered shuttling between nucleus and cytoplasm (Figure 2D). Taken together, we here identify both a rapidly shuttling and a fixed pool of Far1, which supports our model for how compartmentalization is used to generate intergenerational memory.

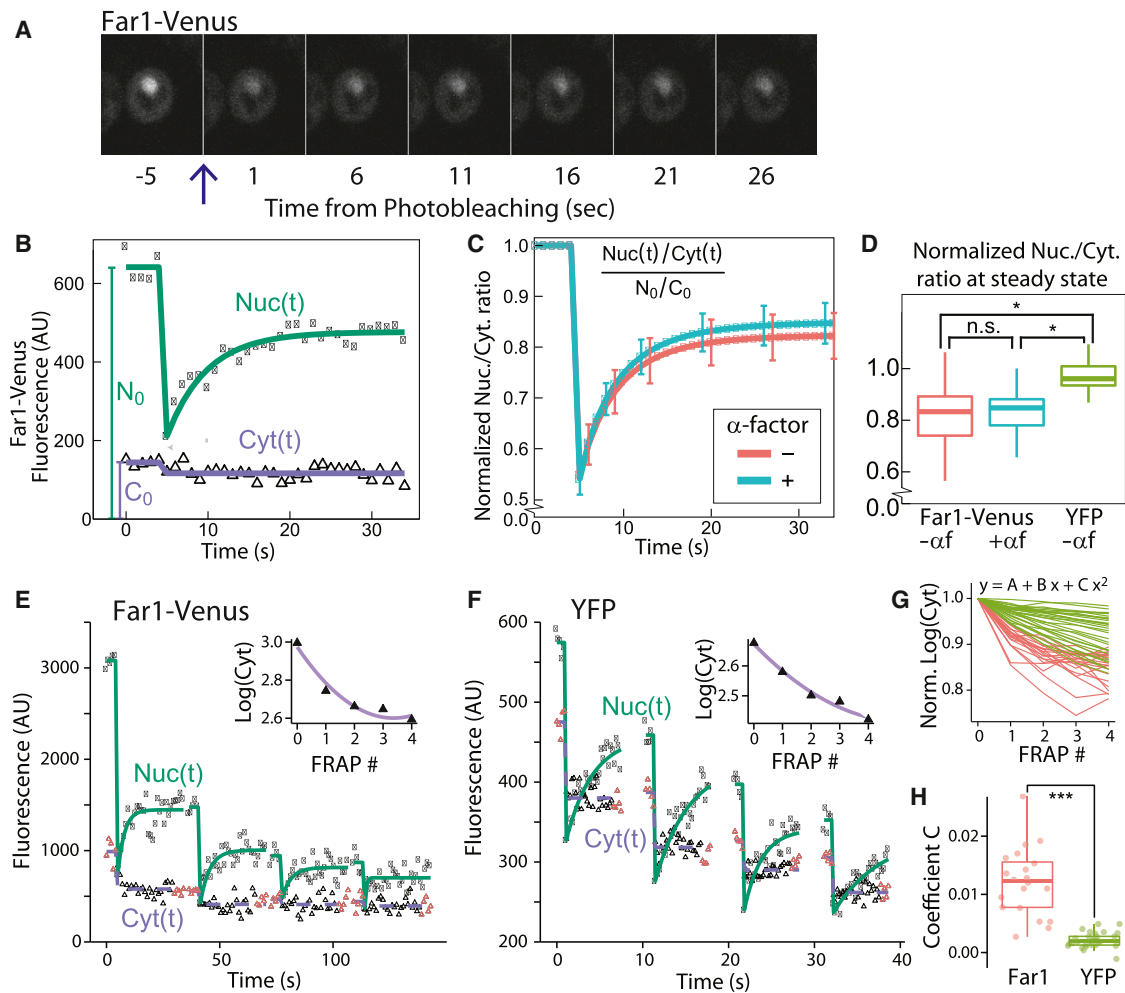
To transmit intergenerational memory, there should be a fixed cytoplasmic pool of Far1. While a single FRAP experiment indicates the presence of a fixed Far1 pool, it does not identify its location. To determine if there is fixed Far1 in the cytoplasm, we photobleached the nucleus four times sequentially and measured depletion of cytoplasmic Far1-*Venus* or YFP (Figures

2E and 2F). In the case of YFP, there is no fixed pool, so that the nuclear-to-cytoplasmic ratio recovers after each photobleaching event. Thus, each bleaching of the nucleus bleaches a constant fraction of the total protein. This leads to a linear relationship between the logarithm of the fluorescence and the number of photobleaching events (see [Supplemental Information](#)). A fixed cytoplasmic pool would result in a deviation from this linear fit. To test for a fixed cytoplasmic pool, we fit the normalized logarithm of the cytoplasmic fluorescence to a quadratic equation for each cell (Figure 2G). We find positive quadratic coefficients for Far1-*Venus* fits indicating the presence of a pool of Far1 that is fixed in the cytoplasm (Figure 2H).

### Pheromone Exposure Is Remembered across the Entire Cell Cycle

To better understand intergenerational memory, we sought to investigate the mechanisms responsible for the increased time to half-maximum concentration of Far1 in 3 nM relative to 0 nM pheromone (Figure 1D). Such an increase could arise due to either increased Far1 synthesis or decreased Far1 degradation, or both. To test for regulated protein degradation, we expressed a *FAR1-Venus* fluorescent fusion protein from a galactose-inducible *GAL1* promoter. We inactivated Far1 synthesis by switching the carbon source from galactose to glucose and measured Far1 half-life post-*Start* (Figure 3A). Far1 stability post-*Start* is only weakly sensitive to pheromone concentration, which suggests that continued synthesis is more likely than increased protein stability to underlie increased inheritance of cytoplasmic Far1 at intermediate pheromone concentrations (Figures 3B and S4A). To test this possibility, we used single molecule fluorescence in situ hybridization (smFISH) (Raj et al., 2008), to measure the amount of *FAR1* mRNA transcripts (Figure 3C). Indeed, *FAR1* transcription was higher in 3 nM compared to 0 nM for cells with small and medium sized buds, corresponding to S and G2 cells respectively (Figures 3D). For large budded cells, likely about to divide, the number of *FAR1* transcripts was similarly high for both conditions, consistent with previous work showing that *FAR1* and other Ste12 transcription factor targets are transcribed at the M/G1 transition, even at 0 nM pheromone (Donic and Skotheim, 2013; McKinney et al., 1993; Oehlen et al., 1996).

To test if the increased Far1 transcription in intermediate pheromone concentrations results from MAPK pathway activity, we examined *STE5-YFP* cells expressing the mating pathway scaffold protein Ste5 fused to a yellow fluorescent protein (Yu et al., 2008). Ste5 localizes to the site of polarized growth when the mating pathway is active (Pryciak and Huntress, 1998; Strickfaden et al., 2007). *STE5-YFP* cells were arrested in 3 nM pheromone and tracked through a cell cycle. The cell perimeter was segmented, linearized, and plotted on a kymograph to visualize the location and intensity of Ste5-YFP on the cell membrane (Figures 3E and 3F). As expected, we observed a transition from a low to a high level of Ste5-YFP at the site of polarized growth upon pheromone arrest (Figure 3G). Upon reentering the cell cycle, Ste5-YFP only partially dissociates from the membrane suggesting that the MAPK pathway remains active through the cell cycle at intermediate pheromone concentrations ( $p < 0.05$  for all comparisons, Figure 3H). We



**Figure 2. A Pool of Far1 Is Fixed in the Cytoplasm**

(A) Fluorescence images from a typical time course, where the nuclear Far1-Venus was photobleached at  $t = 0$ .

(B) Data and model fit for nuclear,  $Nuc(t)$ , and cytoplasmic,  $Cyt(t)$ , Far1.  $N_0$  and  $C_0$  denote the initial nuclear and cytoplasmic fluorescence.

(C) Mean nuclear-to-cytoplasmic ratio of Far1-Venus normalized to its initial value,  $N_0/C_0$ . Bars denote the 95% confidence interval of the mean. We examined pre-Start G1 cells either not exposed to  $\alpha$ -factor (red) or exposed to 500 nM  $\alpha$ -factor (blue).

(D) Distribution of the estimated steady-state value of the normalized Nuc/Cyt ratio after photobleaching. Cells expressing Far1-Venus (blue/red) do not recover the initial nuclear-to-cytoplasmic ratio, while cells expressing the fluorescent protein YFP from an integrated ACT1 promoter (green) recover the initial ratio (see also Figure S3).

(E and F) Nuclear and cytoplasmic fluorescence from Far1-Venus or YFP following four sequential photobleaching events. Inset shows logarithm base 10 of the mean steady-state cytoplasmic fluorescence following the indicated photobleaching event and the associated quadratic fit. Red triangles denote data points used for steady-state estimates.

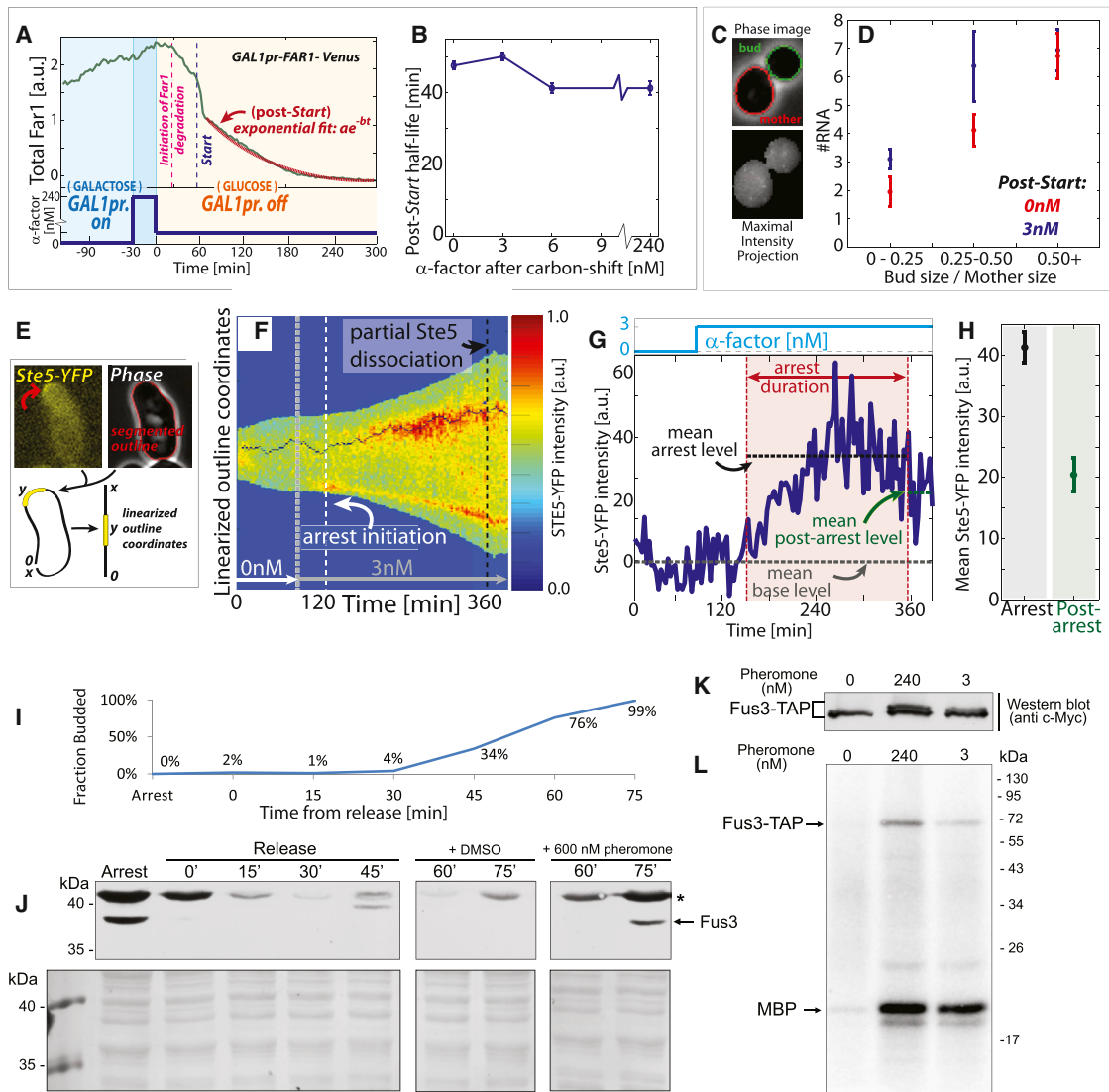
(G) Single cell data for cytoplasmic steady-state fluorescence after normalization to its value prior to the first bleaching event for Far1-Venus (red) and YFP (green).

(H) Distribution of coefficients  $C$  for the quadratic term of the quadratic fit.  $C = 0$  indicates a linear relationship between the logarithm of the cytoplasmic fluorescence and the number of photobleaching events, which corresponds to the case with no fixed cytoplasmic pool.  $C > 0$  indicates the presence of a fixed cytoplasmic pool (see Supplemental Information).

\*denotes  $p < 0.05$ , \*\*\*denotes  $p < 0.001$ , n.s. denotes  $p > 0.05$ . Tukey boxplots in (D) and (H) indicate median, upper, and lower quartiles. Whiskers extend to the most extreme point within  $1.5 \times$  the interquartile range.

also tested if the MAPK Fus3 is active in S/G2/M cells (post-Start) exposed to pheromone as implied by the above results. Fus3 activity correlates with increased nuclear localization and phosphorylation (Blackwell et al., 2003). We therefore measured Fus3 activity using time lapse microscopy and western blot with a phosphospecific antibody (Nagiec and Dohlman, 2012). Consistent with MAPK (Fus3) activity being responsive to pher-

omone concentration in cycling cells, Fus3-GFP nuclear localization quickly decreased in cells in the S/G2/M phases of the cell cycle that experienced a drop in extracellular pheromone concentration (Figures S4B–S4D;  $p < 10^{-4}$ ). Similarly, exposure of cells in the S/G2/M phases of the cell cycle to pheromone increased the amount of phosphorylated Fus3 (Figures 3I–3K). Moreover, we immunoprecipitated Fus3 from S/G2/M cells



**Figure 3. Phormone Exposure Post-Start Is Remembered**

(A) Experiment schematic for measuring stability of Far1 protein post-Start. (B) Post-Start half-life measured after release from phormone arrest in 240 nM to 0, 3, 6 and 240 nM. (C) Example of segmented phase image of mother cell body (red) and bud (green) and their corresponding smFISH maximal intensity projections. Each dot represents a single FAR1 mRNA. (D) Mean number of FAR1 mRNA in cells having small, medium, and large buds. (E) Example segmented phase and Ste5-YFP fluorescence images for a cell arrested in 3 nM  $\alpha$ -factor. Ste5-YFP localizes to the site of polarized growth. (F) Kymograph of example cell in (E). The amount of Ste5 at the site of polarized growth, whose location was determined using the Viterbi algorithm (Forney, 1973). (G) Ste5-YFP trace of example cell shown in (F) indicating levels before, after, and during arrest. (H) Mean Ste5-YFP intensity at the site of polarized growth for pre- and post-Start cells in 3 nM. Membrane fluorescence prior to phormone addition was background subtracted.  $p < 0.05$  for all comparisons. (I-L) Cells were arrested in G1 using phormone and released synchronously through the cell cycle. Fifty minutes after release, after commitment to division, cells were re-exposed to phormone (see methods). (I) Bud index. (J) Top: western blot time course with a phospho-specific antibody indicates presence of phosphorylated Fus3 in S/G2/M cells exposed to phormone; (bottom) Ponceau stained blots are provided as loading controls. (K and L) Fus3-TAP was immunoprecipitated at the 75 min time point, when nearly all cells were in S/G2/M, for cells in 0, 3 or 240 nM phormone. (K) Western blot for this IP indicating increasing Fus3 phosphoshifts in 3 and 240 nM phormone. (L) Fus3 activity on MBP was measured in an in vitro kinase assay using radiolabeling. Error bars in (B), (D) and (H) denote SEM.

exposed to 0, 3 and 240 nM phormone. IP-Fus3 phosphorylated a substrate (MBP) at a rate increasing with phormone concentration (Figure 3L).

Taken together, our data support a model in which cells cycling in intermediate phormone concentrations have increased cytoplasmic Far1 levels due to a partially active

MAPK pathway post-*Start*. Thus, while it is clear that cell-cycle progression inhibits pheromone signaling (Garrenton et al., 2009; Strickfaden et al., 2007; Torres et al., 2011), this inhibition is not complete at intermediate pheromone concentrations. Our data thus shows that intergenerational memory is composed of Far1 accumulated from the entire previous cell division cycle. In other words, cells remember pheromone exposure post-*Start* as well as pre-*Start* from the previous cell division cycle.

### Decreasing Cytoplasmic Far1 Reduces Intergenerational Memory

Our results so far support the model in which an *intergenerational* memory of pheromone exposure is transmitted to newborn daughter cells via stable cytoplasmic Far1. If true, we predict that reducing inherited Far1 by genetic manipulation would result in shorter arrest durations in daughter cells. It was previously shown that deletion of the S-phase cyclins *CLB5* and *CLB6*, but not the G1 cyclins *CLN1* and *CLN2*, resulted in longer arrest durations in 3 nM pheromone (Doncic and Skotheim, 2013) and that ectopic expression of *Clb5* downregulates Far1 (Oehlen et al., 1998). In addition, the S-phase cyclins are nuclear, where Far1 is rapidly degraded (Blondel et al., 2000; Shirayama et al., 1999). We therefore constructed a *CLB5-NES* strain by adding a nuclear export sequence to *CLB5* (Figure 4A). In this strain, the time to half-maximum post-*Start* of cytoplasmic Far1 in 3 nM pheromone was ~45 min, a significant reduction from the ~75 min half-maximum of wild-type cells (Figures 4B and 4C and S5A,B;  $p < 0.01$ ).

In G1, Far1 will be stable because *Clb5* is targeted for degradation by the APC/C following mitosis (Shirayama et al., 1999). Thus, while *CLB5-NES* cells have less cytoplasmic Far1, we expect the smaller amount of inherited Far1 to be just as functional in restraining passage through *Start* as in WT cells. That is, given the same amount of inherited Far1, *CLB5-NES* cells would arrest for similar durations as WT cells. Consistent with these predictions, *CLB5-NES* cells inherited less Far1 and remained arrested for shorter durations relative to WT (Figures 4D, 4E, and S5C). Also as predicted, the relationship between inherited Far1 and arrest duration was statistically similar to WT (Figures 4F and 4G;  $p > 0.1$ ). These data support the interpretation that the *CLB5-NES* allele affects intergenerational memory through a reduction in inherited cytoplasmic Far1 prior to cytokinesis.

### Reducing Inherited Far1 Lowers Mating Efficiency

While our results indicate an intergenerational memory of pheromone exposure from mother to daughter cells, it remains unclear if this intergenerational memory plays a role under other physiological conditions. To test this possibility, we performed a quantitative mating assay using WT and *CLB5-NES* strains. WT cells are able to mate more frequently than *CLB5-NES* cells (Figure 4H,  $p < 0.05$ ). To test that this decrease in mating frequency was not due to a polarization defect we verified that *CLB5-NES* cells polarize similarly to WT cells in presence of a pheromone gradient (Figure S5D). These experiments are consistent with a role for intergenerational memory in physiological conditions.

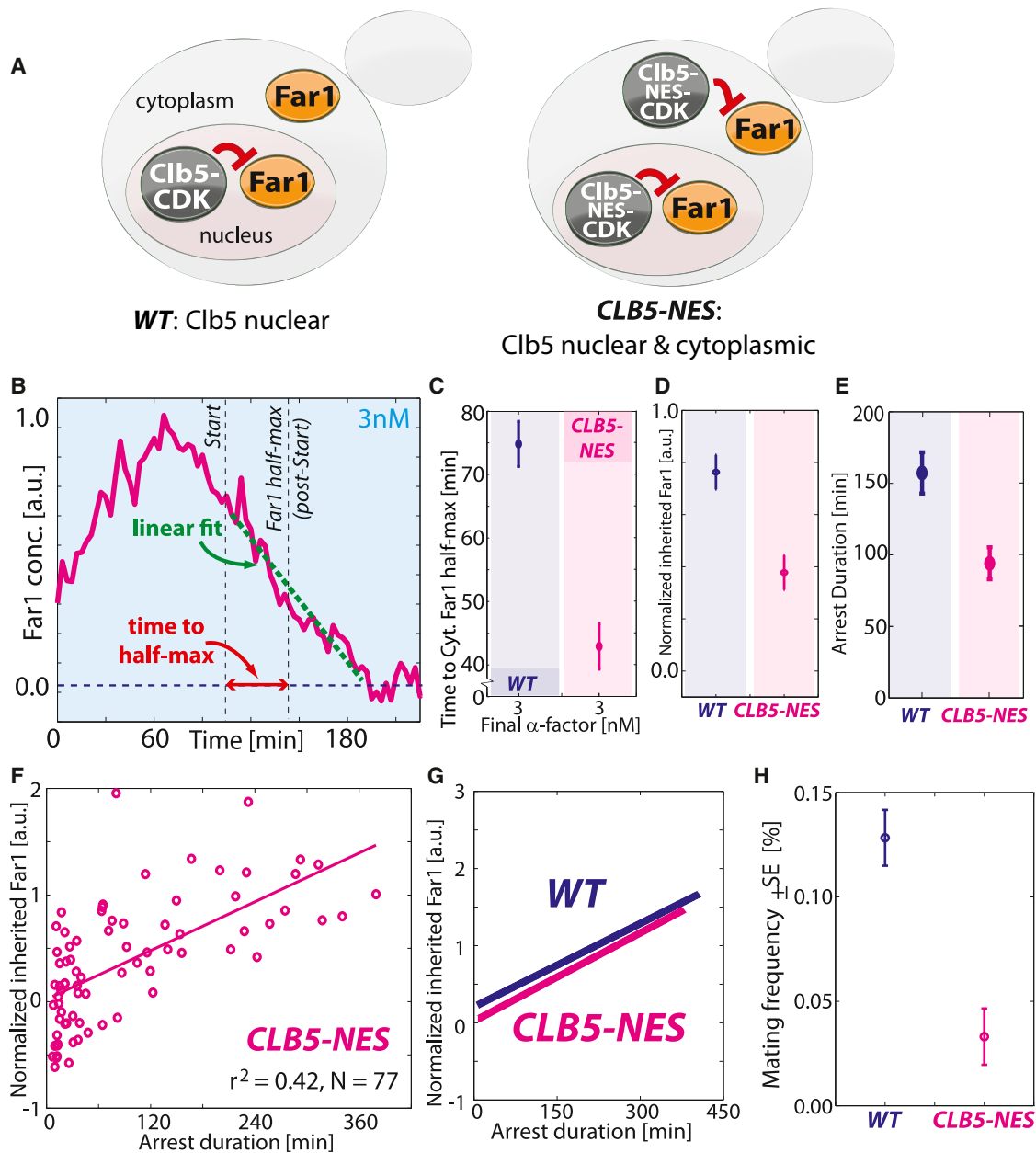
### Far1 Binding to Cdc24 Is Required for Intergenerational Memory

Our results so far identify an intergenerational memory arising from the stability of cytoplasmic Far1. This implies that a non-shuttling cytoplasmic pool of Far1 is inherited to transmit intergenerational memory. Consistent with this model, Far1 has binding partners in the cytoplasm, which we hypothesize serve to anchor Far1. A prime candidate for anchoring is Cdc24, a GTP exchange factor (GEF) regulating cell polarization. Far1 binding to Cdc24 is necessary for pheromone gradient sensing, but not for cell-cycle arrest (Nern and Arkowitz, 1999; Valtz et al., 1995) (Figure 5A). Moreover, Cdc24 is nuclear in G1, but is partially exported to the cytoplasm and plasma membrane during mating arrest in a Far1-dependent manner (Nern and Arkowitz, 2000; Shimada et al., 2000).

To test whether the interaction between Cdc24 and Far1 is required for intergenerational memory, we created strains with the endogenous *FAR1* allele replaced by either *FAR1-D1A* or *FAR1-H7* mutant alleles that express Far1 proteins whose interaction with Cdc24 is greatly reduced (Nern and Arkowitz, 2000; Valtz et al., 1995). As previously reported, both strains arrest in pheromone. However, we identified a slight arrest deficiency and 6 nM pheromone was required to arrest cells for similar durations as WT cells in 3 nM ( $p > 0.05$ ). We therefore used 6 nM for the analysis of *FAR1-D1A* and *FAR1-H7* strains. Consistent with Cdc24 anchoring Far1 in the cytoplasm during arrest, the nuclear fraction of Far1 was increased in *FAR1-D1A* and *FAR1-H7* cells compared to WT cells ( $p < 10^{-3}$ ; Figures 5B, S6A, and S6B). Since nuclear Far1 is rapidly degraded in the cell cycle (Figure 1D), we expected that reduction of cytoplasmic anchoring results in a more rapidly degraded Far1 protein. Indeed, Far1 proteins with reduced Cdc24 interactions reach half-maximum concentration more rapidly following cell-cycle entry (Figures 5C, S6C, and S6D). Finally, we examined the relationship between intergenerational memory and inherited Far1 in *FAR1-D1A* and *FAR1-H7* cells. Consistent with the requirement of a cytoplasmic anchor, and the model that Cdc24 fills this role, post-*Start* Far1 was less stable, less Far1 was inherited, and the intergenerational memory was abolished or greatly reduced in cells expressing Far1 proteins with reduced ability to bind Cdc24 (Figures 5D–5F and S6E–S6G).

To further test the Cdc24 anchoring model, we sought to examine *bni1*  $\Delta$  cells that are unable to export Cdc24 from the nucleus to the shmoo tip during pheromone arrest (Qi and Elion, 2005). *Bni1* is a formin that regulates the polarization of actin cables during mating arrest and is required for cell polarization (Evangelista et al., 1997). We found that *bni1*  $\Delta$  cells arrested as round cells for significant periods of time in G1 when exposed to 6 nM pheromone (Figure S6H). Under these conditions, *bni1*  $\Delta$  cells contained a higher fraction of nuclear Far1, and degraded Far1 more rapidly upon cell-cycle entry compared to WT cells (Figures 5B, 5C, S6B, and S6C). Finally, *bni1*  $\Delta$  cells exhibited no intergenerational memory (Figure 5G). That the localization of Cdc24 outside the nucleus was required for intergenerational memory further supports the role of cytoplasmic Cdc24 as a Far1 anchor.





**Figure 4. Reduction of Cytoplasmic Far1 Decreases Intergenerational Memory**

(A) Clb5 targets Far1 for degradation and is predominantly nuclear in WT cells. Adding a nuclear exclusion sequence (NES) to *CLB5* translocates a fraction to the cytoplasm to target cytoplasmic Far1.

(B) Example time series of Far1 concentration in a *CLB5-NES* cell used to calculate the time to half-maximum post-Start for cytoplasmic Far1.

(C) Mean time to cytoplasmic Far1 half-maximum post-Start for WT and *CLB5-NES* cells first arrested in 240 nM and then released into 3 nM  $\alpha$ -factor.

(D) *CLB5-NES* cells inherit less Far1 than WT and (E) arrest significantly shorter duration ( $p < 10^{-5}$ ).

(F and G) The relationship between the amount of inherited Far1 and the duration of the subsequent arrest is statistically indistinguishable for *CLB5-NES* and WT cells ( $p > 0.05$ ).

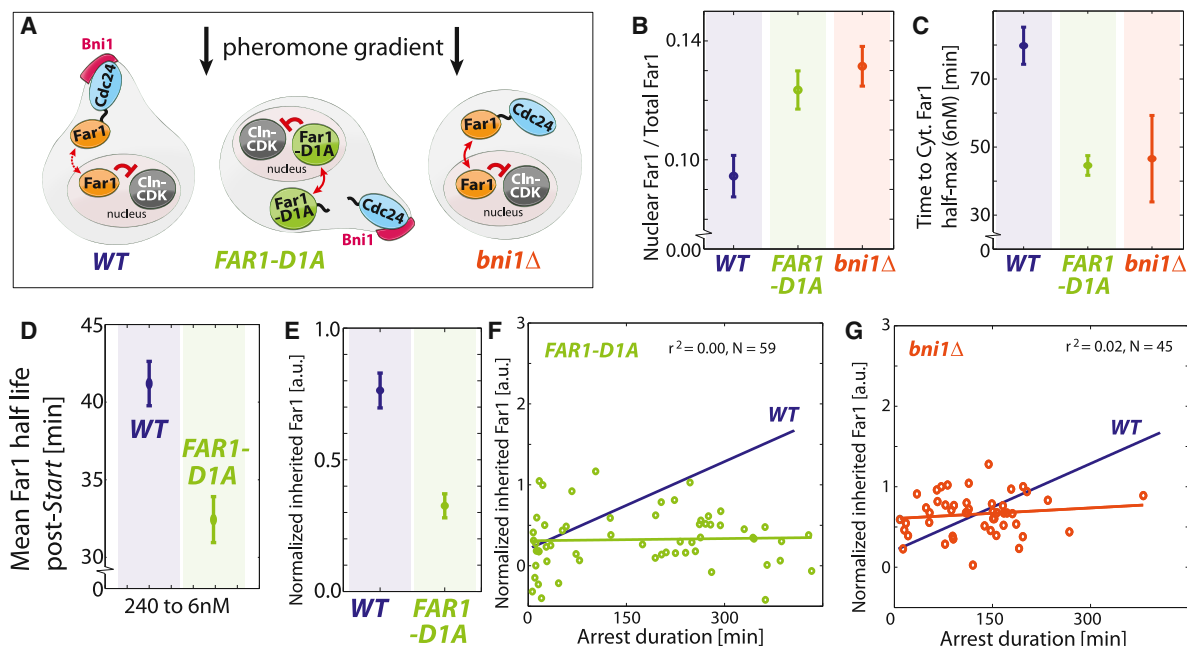
(H) WT and *CLB5-NES* cells exhibit significantly different mating frequencies ( $p < 0.05$ ).

Error bars denote SEM of cells in (C–E) or of replica experiments in (H).

#### Far1 Stability Pre- and Post-Start Is Required for Intra- and Inter-Generational Memory Respectively

The intergenerational memory that we describe here is in addition to the *intragenerational memory* of pheromone exposure en-

coded in Far1 that we previously described (Doncic and Skotheim, 2013). Intragenerational memory allows cells to remember their history of exposure to pheromone during an arrest via the accumulation of Far1. Since Clb5 is targeted for



**Figure 5. Far1 Binding to Cytoplasmic Cdc24 Is Required for Intergenerational Memory**

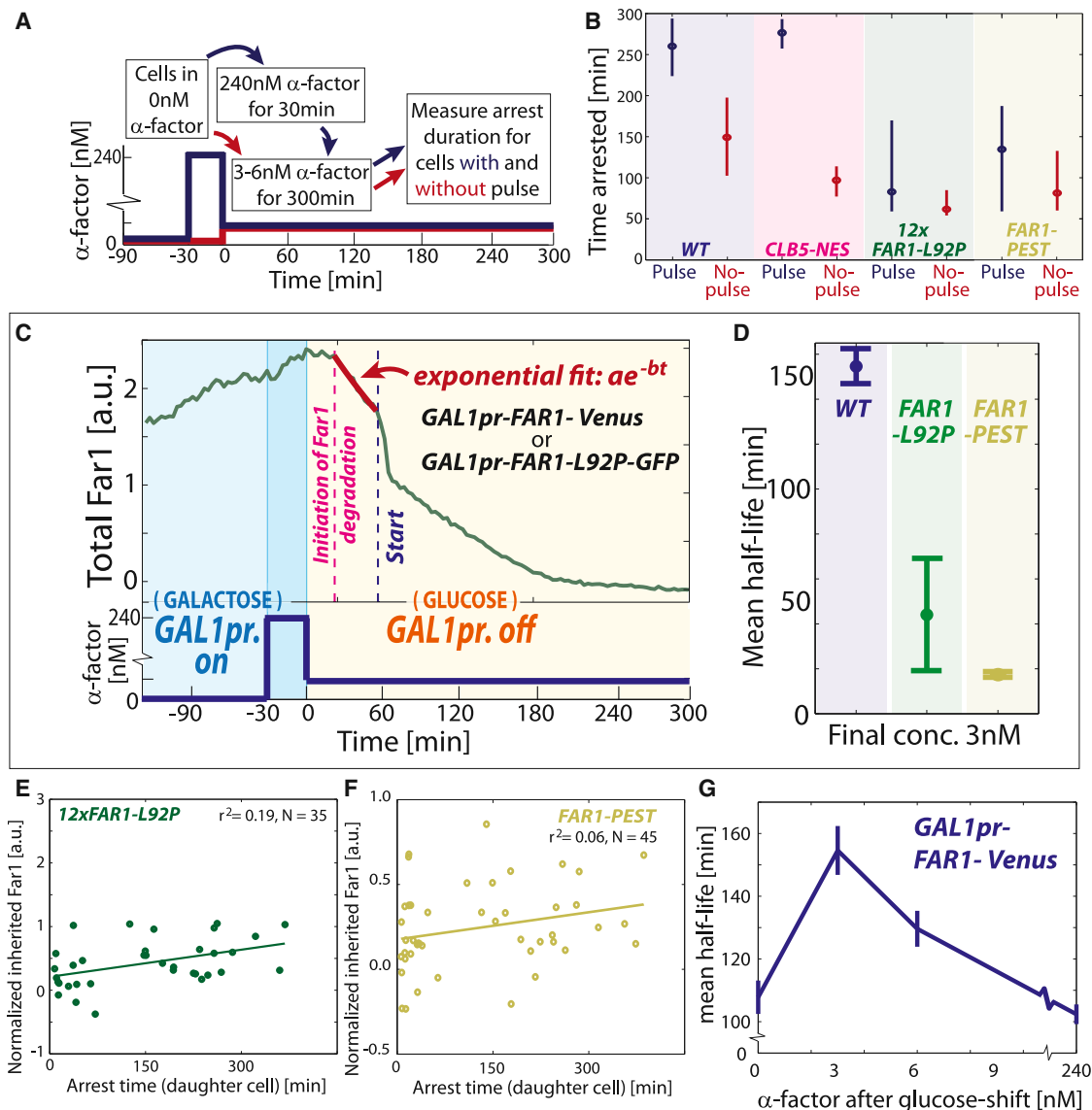
(A) Schematic of the role of Cdc24 with respect to Far1.  
 (B) Fraction of nuclear Far1 in arrested cells.  
 (C) Time to half-maximum Far1 after cell-cycle reentry in 6 nM pheromone.  
 (D) The post-Start stability of Far1-D1A measured as described in Figure 3A.  
 (E) Inherited Far1 in *FAR1-D1A* cells compared to WT ( $p < 10^{-4}$ ).  
 (F and G) No intergenerational memory was observed for *FAR1-D1A* and *bni1Δ* cells.  
 Error bars in (B–E) denote SEM.

degradation in mitosis, and is therefore not active during pheromone arrest, we do not expect cytoplasmic Clb5 to affect intra-generational memory. To test this prediction, we examined cell-cycle progression in cells exposed to different histories of mating pheromone during G1. Cells were either exposed to a brief pulse of high pheromone followed by an intermediate pheromone concentration or just to the intermediate pheromone concentration (Figure 6A). As predicted, both WT and *CLB5-NES* cells experiencing the high pheromone pulse greatly extended arrest duration indicating that while *CLB5-NES* cells have reduced intergenerational memory, their intragenerational memory remains firmly intact (Figures 6B and S7A–S7D). Furthermore, these experiments demonstrate how intergenerational memory is distinct from intragenerational memory and affected by different mutations.

Just as intergenerational memory depends on the stability of Far1 throughout the cell cycle, intragenerational memory should depend on the stability of Far1 during arrest. To destabilize Far1 during pheromone arrest, we generated a *FAR1* allele with the 92<sup>nd</sup> residue mutated from Leucine to Proline (*FAR1-L92P*). This mutation is predicted to generate an additional Cdk consensus phosphorylation site to enhance the degradation of Far1 (E.V. and M.L., unpublished data; Figure S7E). To control for the potentially pleiotropic effects of the L92P mutation, we also generated a *FAR1-PEST* allele, where an otherwise WT *FAR1* allele was fused to the C terminus of *CLN2*, which desta-

bilizes this cyclin (Lanker et al., 1996). To determine the stability of Far1-L92P and Far1-PEST proteins during arrest, we fused them to GFP and expressed them from a *GAL1* promoter. Consistent with these mutations reducing protein stability, the pre-Start half-lives of Far1-L92P and Far1-PEST were reduced to ~50 and ~20 min respectively compared to over 130 min for WT Far1 (Figures 6C and 6D).

To test the dependence of intragenerational memory on Far1 stability, we next constructed a strain containing a single copy of *FAR1-L92P* expressed from its endogenous locus. However, *FAR1-L92P* cells failed to arrest even at high pheromone concentrations. We therefore constructed a strain containing 10–12 copies of *FAR1-L92P* that arrested as WT cells ( $72 \pm 4$  min for WT and  $61 \pm 7$  min for *FAR1-L92P* in 2.7 nM pheromone,  $p = 0.17$ ). To verify that the activity of Far1 remains unaltered in the *FAR1-L92P* strain we also showed that the ability of *12x FAR1-L92P* cells to polarize toward pheromone gradients was similar to WT cells (Figure S7F). Consistent with memory depending on Far1 stability, *12x FAR1-L92P* cells exhibited little if any intragenerational memory despite retaining the ability to arrest at this pheromone concentration (Figures 6B and S7A–S7C). In addition, *12x FAR1-L92P* cells also exhibit no intergenerational memory, most likely because this phenomenon also depends on Far1 stability (Figure 6E). Similarly, the destabilized *FAR1-PEST* strain greatly reduced both intra- and inter-generational memory (Figures 6B, 6F, S7A, and S7G). Taken together, these experiments



**Figure 6. Protein Stability Is Required for Intra- and Inter-Generational Memory**

(A) Experimental schematic for intragenerational memory experiment.

(B) WT and *CLB5-NES* cells have intragenerational memory, where the decision to reenter the cell cycle is based on the history of pheromone exposure during the arrest, while *12x FAR1-L92P* and *FAR1-PEST* cells do not. Medians plotted with 95% confidence intervals computed using 10,000 bootstrap iterations. Note that about half of both the *CLB5-NES* and WT cells exposed to a pulse of high mating pheromone are arrested for the duration of the experiment (Figure S7B). We therefore do not compare arrest durations for WT and *CLB5-NES* cells exposed to a pheromone pulse.

(C) Conditional expression of *FAR1* from a *GAL1* promoter is used to measure half-life pre-Start in a series of pheromone concentrations.

(D) Far1 half-life pre-Start in WT, *FAR1-L92P*, and *FAR1-PEST* cells in 3 nM  $\alpha$ -factor.

(E and F) *12x FAR1-L92P* and *FAR1-PEST* cells lack intergenerational memory as their arrest duration is independent of the amount of inherited Far1.

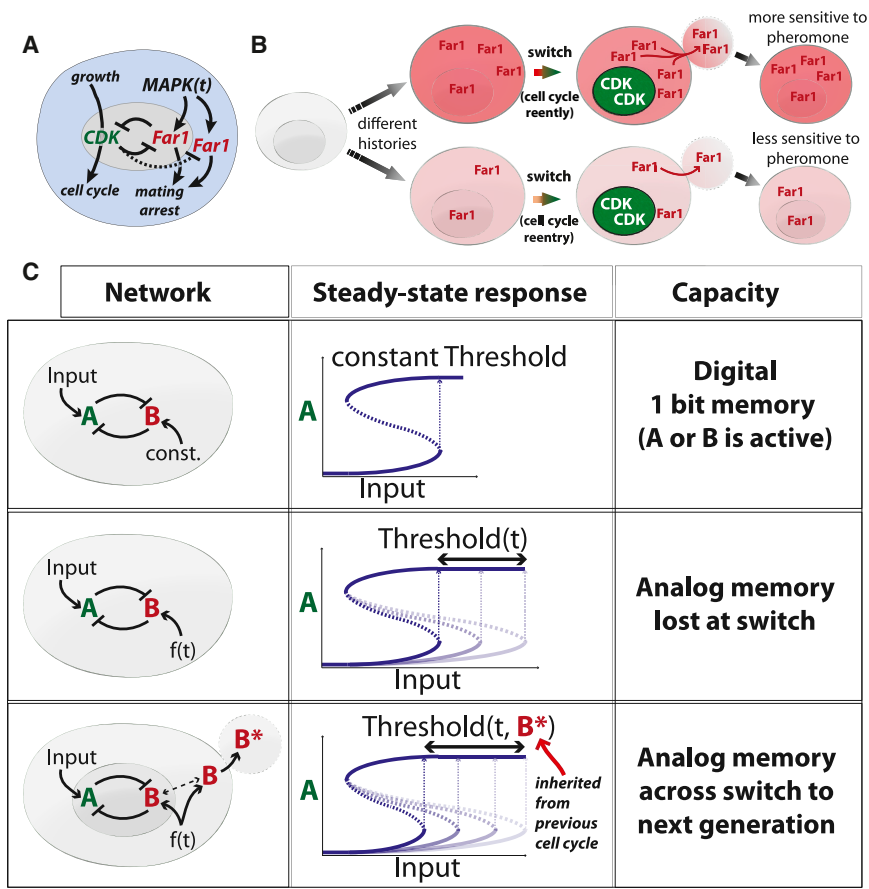
(G) Conditional expression from a *GAL1* promoter is used to measure Far1 half-life  $\pm$  SE pre-Start as in (C), but for a range of pheromone concentrations. Error bars in (D) and (G) denote SEM.

demonstrate the requirement of Far1 stability pre- and post-Start for intra- and inter-generational memory respectively.

### Far1 Stability during Arrest Peaks at Intermediate Pheromone Concentrations

The clear connection between Far1-based memory and protein stability suggested the possibility that WT cells might modulate

Far1 stability to regulate memory. To test this possibility, we measured *FAR1* half-life by expressing it from the *GAL1* promoter and shutting off transcription (Figure 6C). However, we now measure the half-life during pre-Start for cells growing in 0, 3, 6, or 240 nM pheromone. We found that Far1 stability peaked at 3 nM with a  $\sim 150$  min half-life (Figures 6G and S7H and S7I). Far1 stability was reduced to  $\sim 100$  min in



**Figure 7. Compartmentalization Enables Analog Memory to Pass through a Bistable Switch**

(A) The *Start* regulatory network regulating the proliferation-differentiation decision in budding yeast is a bistable switch. (B) Inheritance of cytoplasmic Far1 forms the mechanistic basis of intergenerational memory of pheromone exposure. (C) Top: a double-negative switch with constant signal activating B results in a bistable switch with a single bit of memory, i.e., did the input approach its current level from above or below. Middle: a time-dependent  $f(t)$  signal activates B to allow the time-dependent threshold to encode an analog memory of  $f(t)$ . However, this memory is lost by inactivation upon triggering the double-negative switch. Bottom: compartmentalization allows transmission of analog memory of the time-dependent  $f(t)$  signal across the double-negative (positive) feedback switch.

ferential expression of *CLN3*, the upstream cyclin driving cell-cycle progression in G1 (Laabs et al., 2003). Mother cells produce a burst of *CLN3* expression at the transition from mitosis to G1, while daughter cells do not due to the daughter-specific transcription factors, Ace2 and Ash1 (Di Talia et al., 2009). In addition, differential post-transcriptional regulation of *CLN3* in mother and

0 and 240 nM pheromone. We speculate that this decreased stability might arise from increased G1 cyclin and Fus3 MAPK kinase activities. It is interesting to note that the maximum half-life, i.e., maximal memory, occurs right where the decision to reenter the cell cycle is most sensitive to pheromone concentration.

**DISCUSSION**

To inform our decisions, our experiences are sensed, encoded, and stored as memories. Like us, individual cells also rely on past experience to inform their most important decisions. In budding yeast, one of the most important decisions is whether or not to proliferate or to arrest division and attempt to mate with another haploid cell. Budding yeast invest heavily in this decision, as mutations eliminating the ability to mate provide a ~2% growth rate advantage in pheromone-free conditions (Lang et al., 2009). Perhaps not surprisingly, several distinct types of memory regulate this proliferation-differentiation decision.

The simplest type of memory informing yeast mating is binary and stored as a single bit of information. For example, due to its asymmetric division pattern, a budding yeast cell is either a mother or a daughter. With regards to mating, this bit matters because mother cells are less sensitive to pheromone than daughter cells (Moore, 1984). In part, this is likely due to the dif-

daughter cells may be due to Whi3 (Caudron and Barral, 2013), which both decreases *CLN3* message stability and translation rate (Cai and Futcher, 2013; Gari et al., 2001; Holmes et al., 2013). Yet, all this information pertaining being a mother or daughter cell is binary.

Previously, we identified a continuous, analog form of memory of past pheromone exposure that informs the decision to reenter the cell cycle from pheromone arrest (Doncic and Skotheim, 2013). Cells experiencing higher pheromone concentrations over longer periods of time are more reluctant to reenter the cell cycle. While MAPK pathway activity rapidly responds to reflect the current extracellular pheromone concentration, proteins are more stable so that their level will reflect an integral of past pathway activity (Colman-Lerner et al., 2005; Takahashi and Pryciak, 2008; Yu et al., 2008). More specifically, Far1 accumulates at a monotonically increasing rate with pheromone concentration so that its total amount reflects a combination of arrest duration and pheromone concentration (Chang and Herskowitz, 1990; Doncic and Skotheim, 2013). Thus, the amount of Far1 accumulated during pheromone arrest reflects an integral of pathway activity over time that encodes the history of pheromone exposure into a continuous analog rather than binary variable. However, the mutual inhibition of Far1 and Cdk activities suggested that this analog memory, i.e., the accumulated Far1, would be lost upon flipping the cell-cycle switch (Figures 7A–7C).



Here, we show how the distribution of Far1 into nuclear and cytoplasmic compartments is used to transmit the analog memory of pheromone exposure across the cell cycle switch to the next generation. While nuclear Far1 is rapidly degraded, as expected by the double-negative switch, cytoplasmic Far1 is longer lived and continually synthesized so that a significant amount remains at the end of the cell cycle to be inherited by daughter cells. We show here that this inherited Far1 contributes to increased pheromone sensitivity and thereby allows mother cells to transmit intergenerational memory of pheromone exposure to their daughter cells. We also identified a fraction of fixed cytoplasmic Far1 as key to storing intergenerational memory. If some Far1 were not fixed in the cytoplasm, it would likely be rapidly degraded post-Start due to the high nuclear B-type cyclin activity. While this Far1 fraction is fixed on the minute timescale, we suspect that it is not permanently fixed because nuclear Far1 is important for maintaining the cell-cycle arrest to which the inherited cytoplasmic Far1 eventually contributes. Thus, the slow dissociation of the fixed Far1 is likely central to reading the intergenerational memory of pheromone exposure.

Protein stability is central to both intra- and intergenerational Far1-based analog memory as destabilizing mutations eliminate both. In general, protein stability determines the timescale on which the cell can remember past events. For rapidly degraded proteins, levels will simply reflect the current state of the cell, while for stable proteins, levels will reflect their synthesis over longer periods of time so that their amount can be used to store long-term memories on the timescale of dilution due to cell growth. The demonstrated ability of the cell to regulate protein stability over a wide range of temporal and spatial scales suggests that the analog memory mechanisms discussed here can be easily tuned through mutation and selection.

More broadly, our work illustrates how spatial organization can greatly expand the functionality of signaling motifs. Recently, it has been shown how positive feedback can be enhanced by protein transport within the mammalian mitotic switch (Santos et al., 2012). Activation of Cdk1-Cyclin B complexes within the nucleus recruit additional such complexes to dramatically ramp up nuclear Cdk activity without protein synthesis. However, this represents an enhancement of the well-known ability of positive feedback circuits to generate sharp switches. Here, we have shown how spatial organization allows memory to be transmitted across a positive feedback-driven switch to enable an entirely new and unexpected property of this well-characterized signaling motif. Given the extensive spatial organization within cells, we expect this example to be the first of many in which new signal-processing properties of network motifs are enabled by compartmentalization.

## EXPERIMENTAL PROCEDURES

See additional [Supplemental Information](#) for methods regarding confocal microscopy, FRAP, western blot and kinase assays shown in [Figures 2](#) and [3](#).

### Wide-Field Time Lapse Microscopy and Analysis

A Zeiss Observer Z1 microscope with an automated stage using a plan-apo 63X/1.4NA oil immersion objective and Definite Focus hardware was used to take images every 3 min (6 min for the *FAR1-L92P* strains). We used a CellASIC microfluidics device (<http://www.cellasic.com/>) with Y04C plates. *WHI5-*

*mCherry*, *FAR1-Venus* and *FAR1-GFP* strains were exposed for 750 ms, 300 ms or 150–300 ms using the Colibri 540-80, 505 or 470 LED modules respectively at 25% power. There was no significant photobleaching at our sampling rate ([Figure S7J](#)). *FAR1* activity is not affected by fusion to a fluorescent protein (Doncic and Skotheim, 2013). Image segmentation and quantification was performed as described in (Doncic et al., 2013). We often plot mean values and their associated SE because this gives a graphical representation of statistical significance. Corresponding full distributions can be found in the [Supplemental Information](#).

### Measurement of Inherited Far1

For each cell we determine inherited Far1 to be (Far1-Venus signal – baseline)/(baseline). “Normalized inherited Far1” is the amount at the beginning of G1 above what that cell would be expected to have when cycling in pheromone-free media (see also schematic S21 and [Supplemental Information](#) for details).

### Strains and Media

All strains are congenic with W303 (see [Table S1](#)) and were constructed using standard methods. Yeast were grown in synthetic complete media with 2% glucose unless otherwise stated (2% galactose were used for the Far1 stability experiments in [Figures 3](#) and [6](#)). Before an experiment, cells were grown to an OD < 0.1 after which they were sonicated for ~5 s at 3W intensity. All media were mixed with 20 mg/ml casein (Sigma) to inhibit  $\alpha$ -factor surface adhesion (Colman-Lerner et al., 2005).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.032>.

## AUTHOR CONTRIBUTIONS

A.D. and J.M.S. designed the study. A.D. and O.A. performed all experiments with the following exceptions. E.V. and M.L. identified of the *FAR1-L92P* mutant and performed the Fus3 activity measurements ([Figures 3I–3L](#)). A.G., A.B., G.V., and A.C.-L. performed the FRAP experiments ([Figure 2](#)). A.D., O.A., and J.M.S. wrote the paper.

## ACKNOWLEDGMENTS

Research in the Skotheim laboratory was supported by the Burroughs Wellcome Fund (CASI) and the NIH (GM092925). The Colman-Lerner laboratory was supported by grant PICT2010-2248 from the Argentine Agency of Research and Technology (ANPCyT) and the NIH (GM097479), while the Loog laboratory was supported by Estonian Science Agency Grant IUT2-21.

Received: July 27, 2014  
Revised: November 24, 2014  
Accepted: January 17, 2015  
Published: March 12, 2015

## REFERENCES

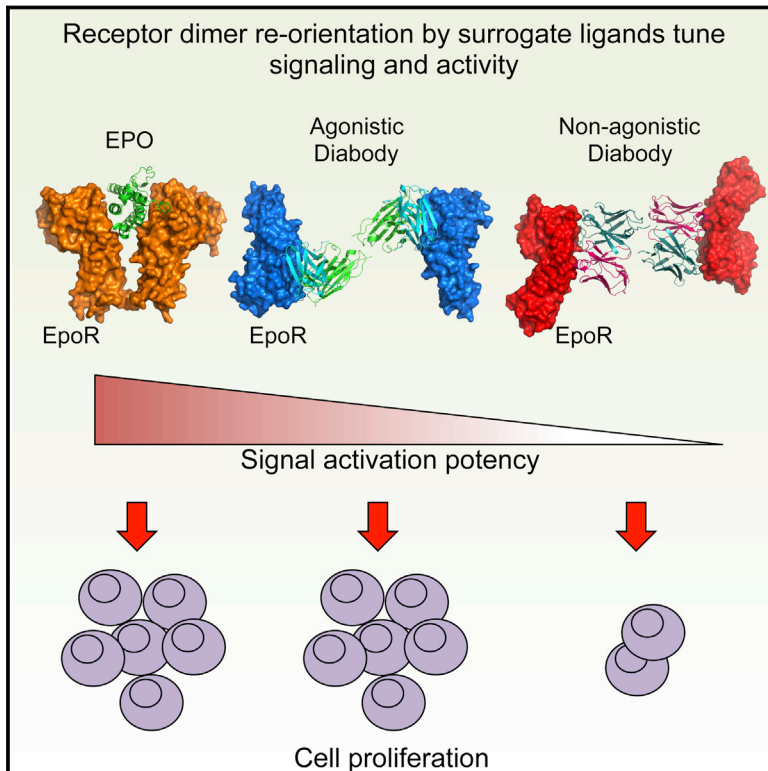
- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.
- Behar, M., Hao, N., Dohlman, H.G., and Elston, T.C. (2008). Dose-to-duration encoding and signaling beyond saturation in intracellular signaling networks. *PLoS Comput. Biol.* 4, e1000197.
- Blackwell, E., Halatek, I.M., Kim, H.J., Ellicott, A.T., Obukhov, A.A., and Stone, D.E. (2003). Effect of the pheromone-responsive G( $\alpha$ ) and phosphatase proteins of *Saccharomyces cerevisiae* on the subcellular localization of the Fus3 mitogen-activated protein kinase. *Mol. Cell. Biol.* 23, 1135–1150.

- Blondel, M., Alepuz, P.M., Huang, L.S., Shaham, S., Ammerer, G., and Peter, M. (1999). Nuclear export of Far1p in response to pheromones requires the export receptor Msn5p/Ste21p. *Genes Dev.* **13**, 2284–2300.
- Blondel, M., Galan, J.M., Chi, Y., Lafourcade, C., Longaretti, C., Deshaies, R.J., and Peter, M. (2000). Nuclear-specific degradation of Far1 is controlled by the localization of the F-box protein Cdc4. *EMBO J.* **19**, 6085–6097.
- Cai, Y., and Futcher, B. (2013). Effects of the yeast RNA-binding protein Whi3 on the half-life and abundance of CLN3 mRNA and other targets. *PLoS ONE* **8**, e84630.
- Caudron, F., and Barral, Y. (2013). A super-assembly of Whi3 encodes memory of deceptive encounters by single cells during yeast courtship. *Cell* **155**, 1244–1257.
- Chang, F., and Herskowitz, I. (1990). Identification of a gene necessary for cell cycle arrest by a negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin, CLN2. *Cell* **63**, 999–1011.
- Chen, R.E., and Thorner, J. (2007). Function and regulation in MAPK signaling pathways: lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1773**, 1311–1340.
- Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C.G., and Brent, R. (2005). Regulated cell-to-cell variation in a cell-fate decision system. *Nature* **437**, 699–706.
- Costanzo, M., Nishikawa, J.L., Tang, X., Millman, J.S., Schub, O., Breitkreuz, K., Dewar, D., Rupes, I., Andrews, B., and Tyers, M. (2004). CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell* **117**, 899–913.
- de Bruin, R.A., McDonald, W.H., Kalashnikova, T.I., Yates, J., 3rd, and Wittenberg, C. (2004). Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell* **117**, 887–898.
- Di Talia, S., Wang, H., Skotheim, J.M., Rosebrock, A.P., Futcher, B., and Cross, F.R. (2009). Daughter-specific transcription factors regulate cell size control in budding yeast. *PLoS Biol.* **7**, e1000221.
- Doncic, A., and Skotheim, J.M. (2013). Feedforward regulation ensures stability and rapid reversibility of a cellular state. *Mol. Cell* **50**, 856–868.
- Doncic, A., Falleur-Fettig, M., and Skotheim, J.M. (2011). Distinct interactions select and maintain a specific cell fate. *Mol. Cell* **43**, 528–539.
- Doncic, A., Eser, U., Atay, O., and Skotheim, J.M. (2013). An algorithm to automate yeast segmentation and tracking. *PLoS ONE* **8**, e57970.
- Errede, B., and Ammerer, G. (1989). STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. *Genes Dev.* **3**, 1349–1361.
- Evangelista, M., Blondell, K., Longtine, M.S., Chow, C.J., Adames, N., Pringle, J.R., Peter, M., and Boone, C. (1997). Bni1p, a yeast formin linking cdc42p and the actin cytoskeleton during polarized morphogenesis. *Science* **276**, 118–122.
- Forney, G.D. (1973). Viterbi Algorithm. *IEEE* **61**, 268–278.
- Gari, E., Volpe, T., Wang, H., Gallego, C., Futcher, B., and Aldea, M. (2001). Whi3 binds the mRNA of the G1 cyclin CLN3 to modulate cell fate in budding yeast. *Genes Dev.* **15**, 2803–2808.
- Garrenton, L.S., Braunwarth, A., Irniger, S., Hurt, E., Künzler, M., and Thorner, J. (2009). Nucleus-specific and cell cycle-regulated degradation of mitogen-activated protein kinase scaffold protein Ste5 contributes to the control of signaling competence. *Mol. Cell Biol.* **29**, 582–601.
- Gartner, A., Jovanović, A., Jeoung, D.I., Bourlat, S., Cross, F.R., and Ammerer, G. (1998). Pheromone-dependent G1 cell cycle arrest requires Far1 phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase, in vivo. *Mol. Cell Biol.* **18**, 3681–3691.
- Hao, N., Nayak, S., Behar, M., Shanks, R.H., Nagiec, M.J., Errede, B., Hasty, J., Elston, T.C., and Dohlman, H.G. (2008). Regulation of cell signaling dynamics by the protein kinase-scaffold Ste5. *Mol. Cell* **30**, 649–656.
- Hartwell, L.H., Culotti, J., Pringle, J.R., and Reid, B.J. (1974). Genetic control of the cell division cycle in yeast. *Science* **183**, 46–51.
- Henchoz, S., Chi, Y., Catarin, B., Herskowitz, I., Deshaies, R.J., and Peter, M. (1997). Phosphorylation- and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev.* **11**, 3046–3060.
- Holmes, K.J., Klass, D.M., Guiney, E.L., and Cyert, M.S. (2013). Whi3, an *S. cerevisiae* RNA-binding protein, is a component of stress granules that regulates levels of its target mRNAs. *PLoS ONE* **8**, e84060.
- Howell, A.S., Jin, M., Wu, C.F., Zyla, T.R., Elston, T.C., and Lew, D.J. (2012). Negative feedback enhances robustness in the yeast polarity establishment circuit. *Cell* **149**, 322–333.
- Jeoung, D.I., Oehlen, L.J., and Cross, F.R. (1998). Cln3-associated kinase activity in *Saccharomyces cerevisiae* is regulated by the mating factor pathway. *Mol. Cell Biol.* **18**, 433–441.
- Kholodenko, B.N., Hancock, J.F., and Kolch, W. (2010). Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* **11**, 414–426.
- Laabs, T.L., Markwardt, D.D., Slattery, M.G., Newcomb, L.L., Stillman, D.J., and Heideman, W. (2003). ACE2 is required for daughter cell-specific G1 delay in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **100**, 10275–10280.
- Lang, G.I., Murray, A.W., and Botstein, D. (2009). The cost of gene expression underlies a fitness trade-off in yeast. *Proc. Natl. Acad. Sci. USA* **106**, 5755–5760.
- Lanker, S., Valdivieso, M.H., and Wittenberg, C. (1996). Rapid degradation of the G1 cyclin Cln2 induced by CDK-dependent phosphorylation. *Science* **271**, 1597–1601.
- Lee, M.J., Ye, A.S., Gardino, A.K., Heijink, A.M., Sorger, P.K., MacBeath, G., and Yaffe, M.B. (2012). Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794.
- Malleshaiah, M.K., Shahrezaei, V., Swain, P.S., and Michnick, S.W. (2010). The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. *Nature* **465**, 101–105.
- McKinney, J.D., Chang, F., Heintz, N., and Cross, F.R. (1993). Negative regulation of FAR1 at the start of the yeast cell cycle. *Genes Dev.* **7**, 833–843.
- Moore, S.A. (1984). Yeast cells recover from mating pheromone alpha factor-induced division arrest by desensitization in the absence of alpha factor destruction. *J. Biol. Chem.* **259**, 1004–1010.
- Nagiec, M.J., and Dohlman, H.G. (2012). Checkpoints in a yeast differentiation pathway coordinate signaling during hyperosmotic stress. *PLoS Genet.* **8**, e1002437.
- Nern, A., and Arkowitz, R.A. (1999). A Cdc24p-Far1p-Gbetagamma protein complex required for yeast orientation during mating. *J. Cell Biol.* **144**, 1187–1202.
- Nern, A., and Arkowitz, R.A. (2000). Nucleocytoplasmic shuttling of the Cdc42p exchange factor Cdc24p. *J. Cell Biol.* **148**, 1115–1122.
- Oehlen, L.J., McKinney, J.D., and Cross, F.R. (1996). Ste12 and Mcm1 regulate cell cycle-dependent transcription of FAR1. *Mol. Cell Biol.* **16**, 2830–2837.
- Oehlen, L.J., Jeoung, D.I., and Cross, F.R. (1998). Cyclin-specific START events and the G1-phase specificity of arrest by mating factor in budding yeast. *Mol. Gen. Genet.* **258**, 183–198.
- Peter, M., and Herskowitz, I. (1994). Direct inhibition of the yeast cyclin-dependent kinase Cdc28-Cln by Far1. *Science* **265**, 1228–1231.
- Peter, M., Gartner, A., Horecka, J., Ammerer, G., and Herskowitz, I. (1993). FAR1 links the signal transduction pathway to the cell cycle machinery in yeast. *Cell* **73**, 747–760.
- Pope, P.A., Bhaduri, S., and Pryciak, P.M. (2014). Regulation of Cyclin-Substrate Docking by a G1 Arrest Signaling Pathway and the Cdk Inhibitor Far1. *Curr. Biol.*
- Pryciak, P.M., and Huntress, F.A. (1998). Membrane recruitment of the kinase cascade scaffold protein Ste5 by the Gbetagamma complex underlies activation of the yeast pheromone response pathway. *Genes Dev.* **12**, 2684–2697.
- Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. *Science* **336**, 1440–1444.
- Qi, M., and Elion, E.A. (2005). Formin-induced actin cables are required for polarized recruitment of the Ste5 scaffold and high level activation of MAPK Fus3. *J. Cell Sci.* **118**, 2837–2848.

- Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879.
- Santos, S.D., Wollman, R., Meyer, T., and Ferrell, J.E., Jr. (2012). Spatial positive feedback at the onset of mitosis. *Cell* 149, 1500–1513.
- Shimada, Y., Gulli, M.P., and Peter, M. (2000). Nuclear sequestration of the exchange factor Cdc24 by Far1 regulates cell polarity during yeast mating. *Nat. Cell Biol.* 2, 117–124.
- Shirayama, M., Tóth, A., Gálová, M., and Nasmyth, K. (1999). APC(Cdc20) promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. *Nature* 402, 203–207.
- Skotheim, J.M., Di Talia, S., Siggia, E.D., and Cross, F.R. (2008). Positive feedback of G1 cyclins ensures coherent cell cycle entry. *Nature* 454, 291–296.
- Strickfaden, S.C., Winters, M.J., Ben-Ari, G., Lamson, R.E., Tyers, M., and Pryciak, P.M. (2007). A mechanism for cell-cycle regulation of MAP kinase signaling in a yeast differentiation pathway. *Cell* 128, 519–531.
- Takahashi, S., and Pryciak, P.M. (2008). Membrane localization of scaffold proteins promotes graded signaling in the yeast MAP kinase cascade. *Curr. Biol.* 18, 1184–1191.
- Thomson, T.M., Benjamin, K.R., Bush, A., Love, T., Pincus, D., Resnekov, O., Yu, R.C., Gordon, A., Colman-Lerner, A., Endy, D., and Brent, R. (2011). Scaffold number in yeast signaling system sets tradeoff between system output and dynamic range. *Proc. Natl. Acad. Sci. USA* 108, 20265–20270.
- Torres, M.P., Clement, S.T., Cappell, S.D., and Dohlman, H.G. (2011). Cell cycle-dependent phosphorylation and ubiquitination of a G protein alpha subunit. *J. Biol. Chem.* 286, 20208–20216.
- Tyers, M., and Fletcher, B. (1993). Far1 and Fus3 link the mating pheromone signal transduction pathway to three G1-phase Cdc28 kinase complexes. *Mol. Cell. Biol.* 13, 5659–5669.
- Valtz, N., Peter, M., and Herskowitz, I. (1995). FAR1 is required for oriented polarization of yeast cells in response to mating pheromones. *J. Cell Biol.* 131, 863–873.
- Yang, Q., and Ferrell, J.E., Jr. (2013). The Cdk1-APC/C cell cycle oscillator circuit functions as a time-delayed, ultrasensitive switch. *Nat. Cell Biol.* 15, 519–525.
- Yosef, N., and Regev, A. (2011). Impulse control: temporal dynamics in gene transcription. *Cell* 144, 886–896.
- Yu, R.C., Pesce, C.G., Colman-Lerner, A., Lok, L., Pincus, D., Serra, E., Holl, M., Benjamin, K., Gordon, A., and Brent, R. (2008). Negative feedback that improves information transmission in yeast signalling. *Nature* 456, 755–761.

# Tuning Cytokine Receptor Signaling by Re-orienting Dimer Geometry with Surrogate Ligands

## Graphical Abstract



## Authors

Ignacio Moraga, Gerlinde Wernig, ..., Jacob Piehler, K. Christopher Garcia

## Correspondence

kcgarcia@stanford.edu

## In Brief

Synthetic ligands called diabodies can change the amplitude and nature of signal activation, or counteract oncogenic ligand-independent intracellular signaling, by re-orienting the geometry of receptor dimerization.

## Highlights

- Ligand-driven re-orientation of receptor dimer topology tunes signaling output
- Diabodies elicit differential signal activation
- Non-agonistic diabodies counteract intracellular oncogenic signaling

## Accession Numbers

4Y5V  
4Y5X  
4Y5Y





# Tuning Cytokine Receptor Signaling by Re-orienting Dimer Geometry with Surrogate Ligands

Ignacio Moraga,<sup>1,2</sup> Gerlinde Wernig,<sup>3,4</sup> Stephan Wilmes,<sup>5</sup> Vitalina Gryshkova,<sup>6</sup> Christian P. Richter,<sup>5</sup> Wan-Jen Hong,<sup>3,7</sup> Rahul Sinha,<sup>3</sup> Feng Guo,<sup>1,2</sup> Hyna Fabionar,<sup>8</sup> Tom S. Wehrman,<sup>9</sup> Peter Krutzik,<sup>9</sup> Samuel Demharter,<sup>10</sup> Isabelle Plo,<sup>11</sup> Irving L. Weissman,<sup>3</sup> Peter Minary,<sup>10</sup> Ravindra Majeti,<sup>3,7</sup> Stefan N. Constantinescu,<sup>6</sup> Jacob Piehler,<sup>5</sup> and K. Christopher Garcia<sup>1,2,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305-5345, USA

<sup>2</sup>Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305-5345, USA

<sup>3</sup>Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA 94305-5345, USA

<sup>4</sup>Department of Pathology, Division of Hematopathology, Stanford University School of Medicine, Stanford, CA 94305-5345, USA

<sup>5</sup>Division of Biophysics, Department of Biology, University of Osnabrück, 49076 Osnabrück, Germany

<sup>6</sup>Ludwig Institute for Cancer Research and de Duve Institute, Université catholique de Louvain, 1200 Brussels, Belgium

<sup>7</sup>Department of Internal Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA 94305-5345, USA

<sup>8</sup>DiscoverRx, 42501 Albrae Street, Fremont, CA 94538, USA

<sup>9</sup>Primity Bio, 3350 Scott Boulevard, Suite 6101, Santa Clara, CA 95054, USA

<sup>10</sup>Department of Computational Biology, Department of Computer Science, University of Oxford, Oxford OX1 3QD, UK

<sup>11</sup>Institut Gustave Roussy, INSERM U1009, 94805 Villejuif, France

\*Correspondence: [kcgarcia@stanford.edu](mailto:kcgarcia@stanford.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.011>

## SUMMARY

Most cell-surface receptors for cytokines and growth factors signal as dimers, but it is unclear whether re-modeling receptor dimer topology is a viable strategy to “tune” signaling output. We utilized diabodies (DA) as surrogate ligands in a prototypical dimeric receptor-ligand system, the cytokine Erythropoietin (EPO) and its receptor (EpoR), to dimerize EpoR ectodomains in non-native architectures. Diabody-induced signaling amplitudes varied from full to minimal agonism, and structures of these DA/EpoR complexes differed in EpoR dimer orientation and proximity. Diabodies also elicited biased or differential activation of signaling pathways and gene expression profiles compared to EPO. Non-signaling diabodies inhibited proliferation of erythroid precursors from patients with a myeloproliferative neoplasm due to a constitutively active JAK2V617F mutation. Thus, intracellular oncogenic mutations causing ligand-independent receptor activation can be counteracted by extracellular ligands that re-orient receptors into inactive dimer topologies. This approach has broad applications for tuning signaling output for many dimeric receptor systems.

## INTRODUCTION

Receptor dimerization is a universal mechanism to initiate signal transduction and is utilized by many growth factors such as cytokines and ligands for receptor tyrosine kinases (RTK), among others (Klemm et al., 1998; Stroud and Wells, 2004; Ullrich and

Schlessinger, 1990; Wang et al., 2009; Wells and de Vos, 1993). Cytokines are a large class of secreted glycoproteins that contribute to regulating the fate and function of most cell types (Bazan, 1990; Liao et al., 2011; Wang et al., 2009). Cytokines bind to the extracellular domains (ECD) of their cell-surface receptors, forming signaling complexes with receptor homo- or hetero-dimers. Although the dimer is the fundamental signaling unit, cytokine receptor-ligand complexes can form in higher-order assemblies (Boulanger et al., 2003; Hansen et al., 2008). In some cases, cytokine receptors may be pre-associated on the cell surface in an inactive state, with the cytokines re-orienting the receptor dimers into an active state (Brooks et al., 2014; Constantinescu et al., 2001; Gent et al., 2002; Livnah et al., 1999). Cytokines such as erythropoietin (EPO) and growth hormone (GH) homodimerize two identical receptor subunits (Constantinescu et al., 1999; Wells and de Vos, 1993), while other cytokines, such as interleukin-2, heterodimerize a shared receptor (common gamma chain) with a cytokine-specific subunit to initiate signaling (Liao et al., 2011; Wang et al., 2009). Cytokine receptor dimerization principally results in activation of intracellular, non-covalently associated Janus kinases (JAKs), which then activate the STAT pathway to modulate gene expression and ultimately determine cell fate (Ihle et al., 1995; O’Shea and Paul, 2010).

Structures of cytokine-receptor ECD complexes from different systems have revealed a diverse range of molecular architectures and receptor dimer topologies that are compatible with signaling (Boulanger et al., 2003; de Vos et al., 1992; Hansen et al., 2008; LaPorte et al., 2008; Livnah et al., 1996; Ring et al., 2012; Syed et al., 1998; Thomas et al., 2011; Walter et al., 1995; Wang et al., 2005). This topological diversity is also apparent for dimeric RTK ECD complexes with their agonist ligands (Kavran et al., 2014; Lemmon and Schlessinger, 2010). Furthermore, monoclonal antibodies, engineered ligands, and other agents that dimerize receptor extracellular domains can have disparate

impacts on signaling, but the topological relationships of these non-native dimers to those induced by the endogenous ligands are unknown (Boersma et al., 2011; Harwerth et al., 1992; Jost et al., 2013; Kai et al., 2008; Li et al., 2013; Müller-Newen et al., 2000; Nakano et al., 2009; Zhang et al., 2012a). Prior studies have shown that cytokine receptor signaling efficiency can be influenced by extracellular domain mutations or structural perturbations (Barclay et al., 2010; Liu et al., 2009; Millot et al., 2004; Rowlinson et al., 2008; Seubert et al., 2003; Staerk et al., 2011). However, the apparent permissiveness in dimer architecture compatible with signaling raises the following questions: to what degree does modulation of receptor-ligand dimer geometry fine-tune receptor activation (Ballinger and Wells, 1998), and could such an approach constitute a practical strategy to control dimeric receptor signaling output? Correlating the structure of a receptor-ligand complex in different dimerization topologies to functional properties, including membrane-proximal and membrane-distal signaling outputs would be informative in addressing this question.

On one hand, prior studies showing that cytokine-induced intracellular signaling could be activated through chimeric receptors containing alternative ECDs demonstrated that constraints on dimerization geometries compatible with signaling were loose to some degree (Heller et al., 2012; Ohashi et al., 1994; Pattyn et al., 1999; Socolovsky et al., 1998). On the other hand, a series of studies comparing activation of EpoR by its natural ligand EPO versus synthetic peptides concluded that small changes in dimer orientation could modulate signal strength (Livnah et al., 1996, 1998; Syed et al., 1998). However, these studies left open the question of whether the observed signaling efficiency differences were attributable to alternative dimer topologies or ligand affinity. In one example, it was reported that an EPO agonist peptide (EMP-1) could be converted into a non-activating, or “antagonist” peptide (EMP-33) through a chemical modification (Bromination) of the EMP-1 peptide. Crystal structures of both peptide ligands bound to the extracellular domains of EpoR revealed dimeric complexes (Livnah et al., 1996, 1998); however, it was noted that the non-signaling EMP-33/EpoR ectodomain dimer angle differed by an  $\sim 15^\circ$  rotation versus the agonist EMP-1/EpoR dimeric complex (Livnah et al., 1998; and Figure S1A). The lack of signal initiation by the EMP-33 peptide was attributed to this small change in the EpoR ECD dimer angle.

## RESULTS

### EPO Receptor Dimerization and Signal Activation Induced by EMP Peptides

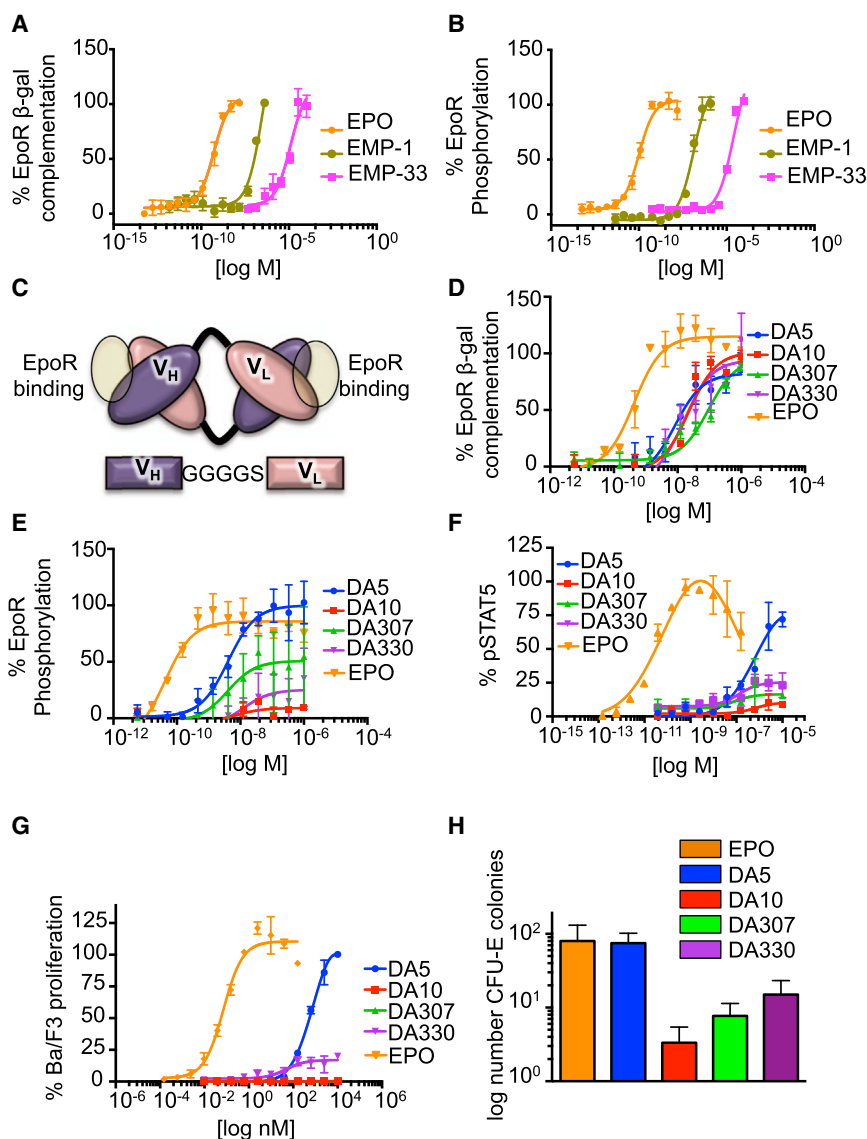
Given the diverse range of dimer topologies evident in agonistic cytokine-receptor complexes (Wang et al., 2009), that in many cases exceed  $15^\circ$  angular differences, we revisited the striking observation seen with the EPO peptide ligands. We explored the biological activity of these peptides using EpoR reporter cells we developed that gave us the ability to test EpoR signaling by receptor phosphorylation but, importantly, also using a beta-galactosidase complementation system that is a sensitive reporter of EPO-induced EpoR oligomerization in physiologic conditions at  $37^\circ\text{C}$ , which directly informs on early signaling

and internalization (Wehrman et al., 2007). First, we synthesized the EMP-1 and EMP-33 peptides and found that EMP-1 binds EpoR with a  $K_D$  of 1  $\mu\text{M}$ , whereas EMP-33 binds EpoR with a  $K_D$  of more than 50  $\mu\text{M}$  (Figures S1B and S1C). The low affinity of EMP-33 prompted us to ask whether its lack of receptor activation is due to low occupancy of the receptor on the cell. We measured the actions of both peptides at inducing signaling and receptor dimerization on cells at a wide range of concentrations. At 10  $\mu\text{M}$  of peptide, only EMP-1 induced dimerization and phosphorylation of EpoR at levels comparable to those achieved by EPO stimulation (Figures 1A and 1B). At higher concentrations of peptide (100  $\mu\text{M}$ ), approaching that used for co-crystallization of both the agonistic and non-signaling dimeric EpoR/peptide complexes, EMP-33 induced a similar degree of receptor dimerization and phosphorylation of EpoR as EMP-1 and EPO itself (Figures 1A and 1B). Thus, when EMP-33 is applied at concentrations that dimerize EpoR on cells, the dimer geometry of the EMP-33/EpoR complex is competent to initiate signaling. The different signaling potencies exhibited by the EPO mimetic peptides appear to be primarily due to their relative EpoR binding affinities.

### EpoR Diabodies Induce Different Degrees of Agonism Activity

We turned our attention to developing surrogate cytokine ligands that could induce much larger topological differences in the EpoR dimer and enable a systematic study relating dimer architecture to signaling and function. We reasoned that diabodies, which are covalently linked dimeric antibody  $V_H/V_L$  variable domain fragments (Fvs) possessing two binding sites, could dimerize and possibly induce signaling of the EpoR, albeit at significantly larger inter-dimer distances than induced by EPO. Additionally, diabodies might be constrained enough to allow crystallization of their complexes with EpoR so that we can directly visualize the dimeric topologies (Perisic et al., 1994). By comparison, whole antibodies have been shown to activate cytokine receptor signaling in many systems, presumably by dimerization (Müller-Newen et al., 2000; Zhang et al., 2012a, 2013). However, the segmental flexibility of intact antibodies has precluded a structural analysis of intact dimeric agonist complexes that can be related to the biological activities.

We synthesized genes of four previously reported anti-EpoR antibodies (Lim et al., 2010) and re-formatted their  $V_H$  and  $V_L$  domains into diabodies (Figure 1C). The four diabodies bound EpoR with approximately similar affinities (Figure S2) and multimerized EpoR with similar efficiency (as measured by  $\text{EC}_{50}$ ), albeit less efficiently than EPO (Figure 1D). However, they induced EpoR phosphorylation with very different relative efficiencies, ranging from full agonism (DA5) to very weak partial agonism (DA10) (Figure 1E). The four diabodies also exhibited different extents of STAT5 phosphorylation (Figure 1F), STAT5 transcriptional activity (Figure S3A), Ba/F3 cell proliferation (Figure 1G), and CFU (colony forming unit)-E colony formation (Figure 1H). These dramatic differences in diabody-induced signaling and functional activities persist at saturating ligand concentrations, so are not attributable to significantly different relative affinities for EpoR or to a stronger EpoR internalization induced by the weak agonist diabodies (DA10, DA307, and



**Figure 1. EpoR Dimerization and Signaling Potencies Induced by EMPs and Diabodies**

(A and B) Levels of EpoR dimerization (A) and phosphorylation (B) promoted by EMPs at the indicated doses. Data (mean  $\pm$  SD) are from four independent replicates.

(C) Schematic view of a bivalent diabody molecule.  $V_H$  is connected to the  $V_L$  domain by a short Gly-linker. EpoR binding sites in the diabody are highlighted with a yellow circle.

(D and E) Levels of EpoR dimerization (D) and phosphorylation (E) promoted by diabodies at the indicated doses. Data (mean  $\pm$  SD) are from four independent replicates.

(F) Percentage of pSTAT5 activation induced by the indicated doses of EPO or the four diabodies in Ba/F3 EpoR cells. Data (mean  $\pm$  SD) are from two independent experiments.

(G) Ba/F3 proliferation in response to EPO or the four diabodies. Data (mean  $\pm$  SD) are from two independent replicates.

(H) Number of CFU-E colonies derived from mouse bone marrow induced by EPO and the four diabodies. Data (mean  $\pm$  SD) are from three different experiments.

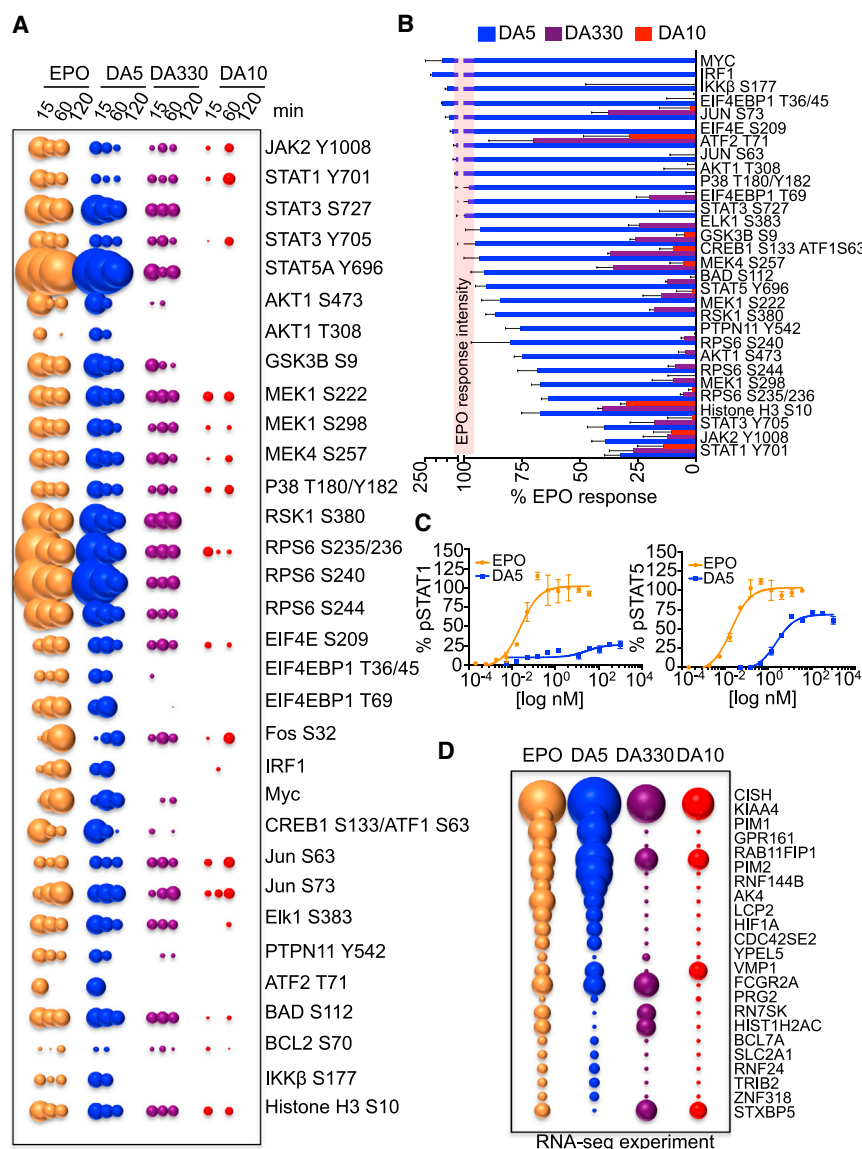
See also Figures S1, S2, and S3.

DA330) because the internalization closely correlated with their signaling efficacies (Figure S3B).

### EpoR Diabodies Induce Differential Signal Activation

STAT5 is the most prominent STAT protein activated by EPO (Constantinescu et al., 1999). However additional signaling pathways, including other STATs (STAT1 and STAT3), the MAPK pathway, and the PI3K pathway, are also activated by this cytokine and fine-tune its responses (Constantinescu et al., 1999). “Biased” signal activation is a phenomenon that has been described for G-protein-coupled receptor (GPCR) ligands, where one GPCR can differentially activate signaling pathways (e.g., beta-arrestin versus G protein), depending on the ligand (Drake et al., 2008). Thus, we asked whether similar differential signal activation could be observed in a dimeric single-pass transmembrane receptor such as the EPO-EpoR system. We studied the activation of 78 different signaling molecules (Table S1) by

phospho-flow cytometry in the EPO-responsive cell line UT7-EpoR. EPO and the diabodies induced the activation of 33 signaling proteins, including members of the STAT family (STAT1, STAT3, and STAT5), MAP kinase family (MEK and p38), and PI3K family (Akt, RSK1, and RPS6) (Figure 2A). We also observed the upregulation of known EPO-induced transcription factors such as Myc, cFos, IRF1, and Elk (Figure 2A). In agreement with our previous results, the signaling potencies exhibited by the three diabodies ranged from full agonism for DA5 to partial agonism for DA330 and non-agonism for DA10 (Figure 2A). Interestingly, the diabodies did not activate all 33 signaling molecules to the same extent (Figure 2B). When the signal activation levels induced by the three diabodies after 15 min stimulation were normalized to those induced by EPO, we observed that, although EPO and DA5 induced similar levels of activation in the majority of the signaling pathways analyzed, DA5 activated some of them to a lower extent than EPO (Figure 2B). Among those, STAT1 and STAT3 activation were the most affected, with DA5 inducing 30% of the STAT1 and 40% of the STAT3 activation levels induced by EPO (Figure 2B). Interestingly, STAT3 S727 phosphorylation, which requires MAPK activation (Decker and Kovarik, 2000), was equally induced by EPO and DA5, which is consistent with the two ligands activating the MAPK pathway to the same extent (Figures 2A and 2B). Dose/response studies in UT7-EpoR cells confirmed these observations and showed that DA5 activates STAT1 to a lesser extent than EPO, while still promoting comparable levels of STAT5



**Figure 2. “Biased” Signaling Activation Induced by the Diabodies**

(A) Bubble plot representation of the signaling pathways activated by EPO and the three diabodies at the indicated times in UT-7-EpoR cells. The size of the bubble represents the intensity of the signal activated.

(B) The levels of signal activation induced by the three diabodies at 15 min of stimulation were normalized to those induced by EPO and order based on signaling potency. The red line represents the EPO signaling activation potency normalized to 100%. Data (mean  $\pm$  SD) are from three independent replicates.

(C) pSTAT1 and pSTAT5 dose-response experiments performed in UT-7-EpoR cells stimulated with EPO or DA5 for 15 min. Data (mean  $\pm$  SD) are from two independent replicates.

(D) Bubble plot representation of genes induced by EPO and the three diabodies after stimulation of MEP cells for 2 hr. The size of the bubble represents the fold of gene induction.

See also Figure S3 and Table S1.

activation (Figure 2C). Thus, biased signaling can be induced through the dimeric EpoR with surrogate ligands.

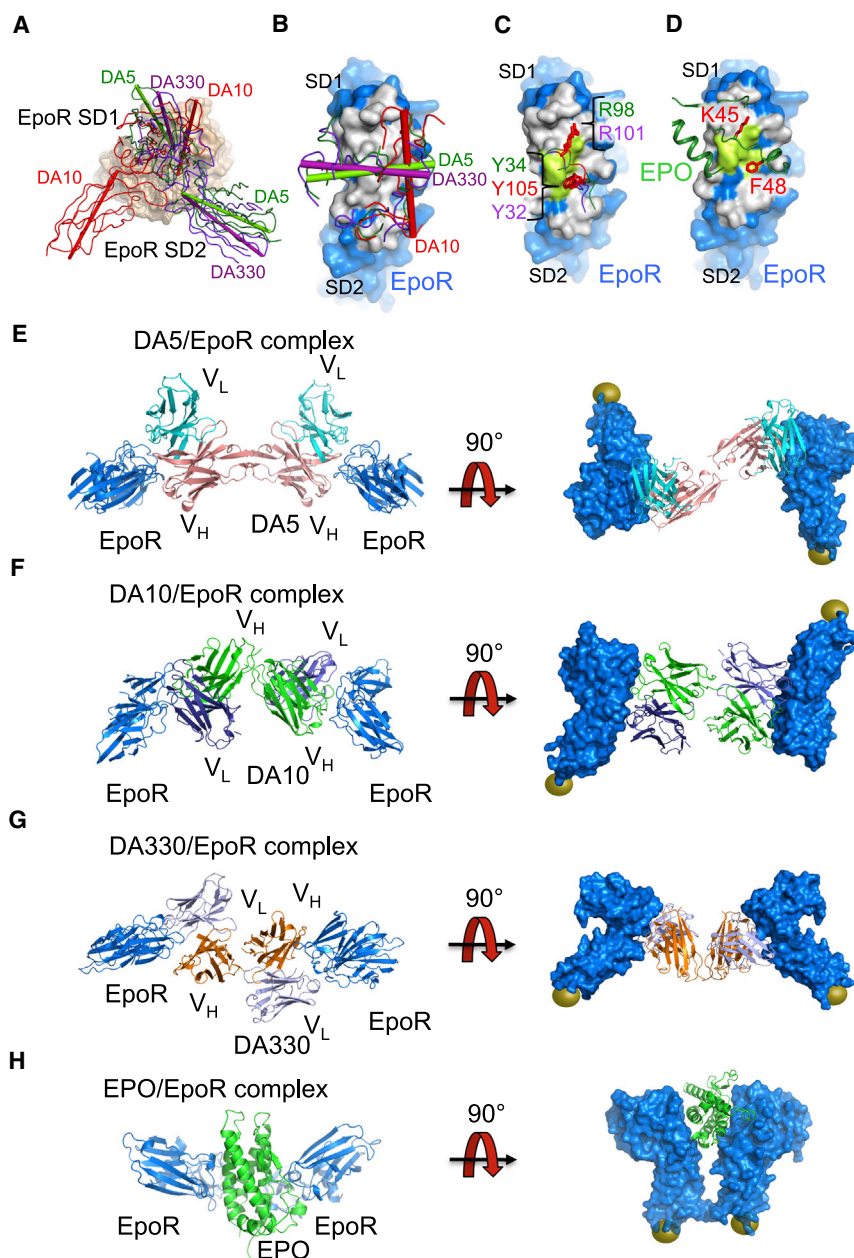
Next, we studied how different signal activation amplitudes exhibited by the diabodies at the membrane-proximal level would impact their membrane-distal gene expression programs. We carried out RNA sequencing (RNA-seq) studies of EPO-responsive genes in purified human primary megakaryocytoerythroid progenitor (MEP) cells derived from bone marrow of a normal subject (Figure 2D). Temporally, MEPs are the first progenitors to robustly express EpoR during hematopoiesis in humans (Seita et al., 2012). In agreement with the signaling data, the relative gene-induction potencies exhibited by the diabodies matched their signaling efficacies (i.e., DA5 > DA330 > DA10) (Figure 2D). DA5 induced a very similar gene induction profile to EPO but with some differences, with a small subset of genes (e.g., *Pim2* and *RN7SK*) being differentially regulated by DA5 when compared to EPO.

ferences likely result from the use of different cell types in the two assays. Overall, our signaling and gene expression data show that the diabodies exhibit various degrees of differential signaling properties relative to EPO and to one another.

### Alternative EpoR Dimer Orientation and Proximity Result in Different Degrees of Agonism

To explore the structural basis for the differential signaling activation exhibited by the diabodies, we expressed and purified three diabody/EpoR complexes (DA5, DA10, and DA330) from baculovirus-infected insect cells. All exhibited molecular weights of 97–98 kDa as measured by multi-angle light scattering (MALS) chromatography, in agreement with a 2:1 complex stoichiometry (two EpoR bound to one diabody [Figure S4A]). We crystallized the diabody/EpoR complexes (DA5 [2.6 Å], DA10 [3.15 Å], and DA330 [2.85 Å]) and determined their structures by molecular replacement (Figure 3 and Table S2). The diabody subunit





**Figure 3. Crystal Structures of DA5, DA10, and DA330 in Complex with EpoR**

(A) Overlay of the three diabody-EpoR complexes. EpoR binding to DA5 is colored green, EpoR binding to DA10 is colored red, and EpoR binding to DA330 is colored purple. The DA330 crystal lattice appears to contain domain-swapped diabodies as scFv in similar but not identical subunit relationships.

(B) Diabodies binding footprint on the EpoR surface. Amino acids on EpoR interacting with the diabodies are colored white. DA5 CDRs are colored green; DA10 CDRs are colored red, and DA330 CDRs are colored purple. Vectors connecting the V<sub>H</sub> CDR1 and the V<sub>L</sub> CDR1 in the diabodies define the binding topology of the three diabodies-EpoR complexes.

(C and D) Diabodies and EPO binding footprint on the EpoR surface. Hotspot interactions on EpoR are colored lime and are shared by the diabodies and EPO. Diabodies use Y34, R98 (DA5), Y105 (DA10), Y32, and R101 (DA330) to interact with the amino acids forming the two hotspots on EpoR. EPO uses similar chemistry with F43 and K45 filling the two hotspot pockets on EpoR.

(E-H) Crystal structures of DA5 (E), DA10 (F), DA330 (G), and EPO (H) dimerizing two EpoR are shown in top (left) and side (right) views. In the side view representation, EpoR is depicted as surface. Yellow spheres represent the C-terminal region of the SD2 EpoR domain.

See also Figures S4, S5, and S6 and Table S2.

relationships are clear for the most and least potent diabody complexes (DA5 and DA10, respectively). For the DA330/EpoR complex, the crystal appears to contain domain-swapped diabodies as “back-to-back,” single-chain Fvs that pack in similar, but not identical, subunit relationships as diabodies. The MALS data show that all of the diabodies are the expected 2:1 complexes in solution.

All three diabodies converge on the protruding “elbow” of EpoR that also serves as the EPO binding site (Figures 3, S4, and S5). When the diabody V<sub>H</sub>/V<sub>L</sub> modules are aligned, the EpoR’s “rotational” binding topology is most similar between DA5 and DA330, with DA10 being markedly different (Figure 3A). Although DA5 and DA330 both bind horizontally and differ pri-

marily in their vertical “tilt” (~14°), DA10 is orthogonally disposed relative to the other two (Figure 3B). In a striking example of chemical mimicry of EPO binding, the diabody CDR loops use two patches of basic (Arg98/Arg101 of DA5 and DA330, respectively) and hydrophobic (Tyr34/Tyr105/Tyr32 of DA5/DA10/DA330, respectively) residues in a nearly identical manner as residues presented on the EPO helices (Lys45 and Phe48) in the EPO site I binding interface to engage the same regions of the EpoR binding site (Figures 3C and 3D).

The overall architectures of the three diabody/EpoR complexes (Figures 3E–3G) are quite distinct from that of the EPO/EpoR complex, which dimerizes two molecules of EpoR in a classical Y-fork cytokine-receptor architecture, resulting in close proximity between the C termini of the membrane-proximal EpoR ECDs (Figure 3H). In contrast, the diabodies impose much larger separation between the two EpoR molecules with distances ranging from ~127 Å in the case of the DA5/EpoR (full agonist) complex to ~148 Å, as in the case of the DA10/EpoR complex (non-agonist) (Figures 3E–3G). The exact EpoR dimer separation is uncertain for the partial agonist DA330 due to the domain swapping. Interestingly, the relative EpoR dimer distances observed in the full and non-agonist diabody/EpoR

complexes correlate with their signaling potencies in that the full agonist DA5 dimer is closer together, whereas the non-agonist DA10 is further. One caveat is that the diabody molecules themselves are not rigid—they exhibit flexibility in the linker and hinge angles relating the two  $V_H/V_L$  modules, raising the question of whether we captured one of a range of dimer angles that could be enforced by crystal lattice contacts. We performed conformational sampling studies exploring the relationship between the EpoR separation distance as a function of the diabody hinge angle on the full agonist DA5 and the non-signaling DA10 (Figure S6 and Movies S1 and S2). The results of these studies show that the thermodynamically permitted variation in diabody hinge angles appears to occupy a few energy minima, leading to only a small range of alternative conformations (i.e., distances) around that seen in the crystal structures (Figure S6 and Movies S1 and S2). The sampling of these alternative conformations has minor consequences on the inter-EpoR distances.

It is important to emphasize that, because we observe differences in both the EpoR/diabody docking angles (Figures 3A and 3B) and the distances between EpoR C termini in the dimeric complexes, we cannot say whether distance or geometry/topology, or a combination of both factors, is responsible for the differences in signaling between the complexes. However, that the differences in signaling amplitude correlate with alternative overall extracellular dimer topologies appears quite clear. Such large differences in extracellular architecture would likely influence the relative orientation and proximity of the two JAKs associated with the membrane proximal intracellular domains of the receptors and impact their subsequent phosphorylation profiles (Figure 5A).

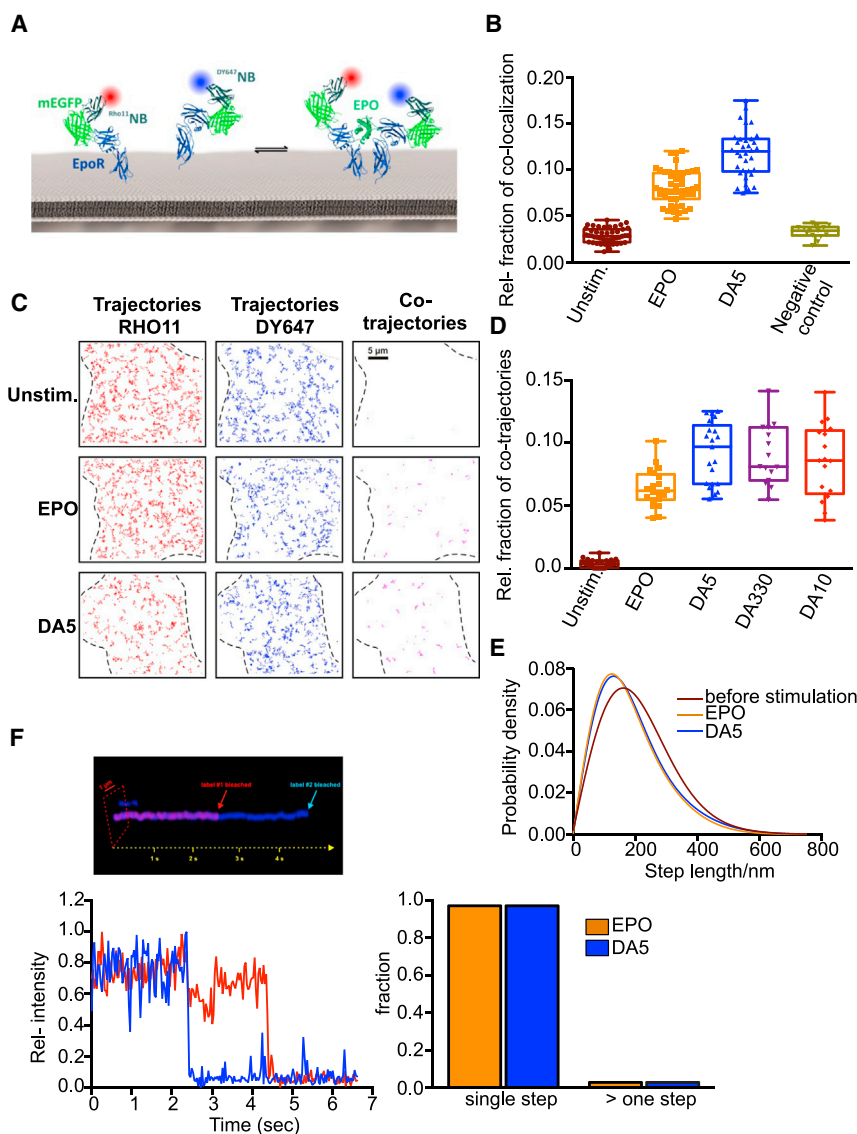
### Comparable Spatiotemporal Dynamics of EpoR Assembly by EPO and Diabodies

An important mechanistic question is whether the diabody/EpoR complexes on the cell surface are indeed homodimers or higher-order species due to clustering of preformed EpoR dimers, which has been reported previously (Constantinescu et al., 2001; Livnah et al., 1999). To explore the ability of the diabodies to dimerize EpoR in the plasma membrane, we probed the assembly and diffusion dynamics of signaling complexes by dual-color single-molecule imaging. For this purpose, EpoR fused to an N-terminal monomeric EGFP (mEGFP) was expressed in HeLa cells and labeled by addition of anti-GFP nanobodies (Rothbauer et al., 2008), which were site-specifically conjugated with DY647 and ATTO Rho11, respectively (Figure 4A). We labeled the receptors extracellularly so as not to introduce fusion proteins to the intracellular regions that may result in artifactual dimerization behavior. Efficient dual-color labeling suitable for long-term observation of individual EpoR was achieved with typical densities of  $\sim 0.3$  molecules/ $\mu\text{m}^2$  in both channels, which was exploited for co-localization and co-tracking analysis. In the absence of an agonist, independent diffusion of EpoR molecules could be observed (Movie S3) with no significant single-molecule co-localization beyond the statistical background (Figure 4B). Single-molecule co-tracking analysis corroborated the absence of pre-dimerized EpoR at the plasma membrane (Figures 4C and 4D). Upon addition of EPO, dimerization of EpoR was detectable by both co-localization and co-tracking

analysis (Movie S3 and Figures 4B–4D). Individual receptor dimers could be tracked (Movie S3), and a clear decrease in their mobility compared to EpoR in absence of ligand was identified (Figure 4E). Stimulation of EpoR endocytosis in presence of EPO was observed, which was accompanied by an increased fraction of immobile EpoR molecules in presence of EPO. The stoichiometry within individual complexes was analyzed by photobleaching at elevated laser power. Single-step photobleaching confirmed the formation of EpoR dimers in the plasma membrane (Movie S4 and Figure 4F). Upon labeling the mEGFP-EpoR only with ATTO-Rho11, two-step bleaching could be observed only in presence of EPO (Movie S5). For all diabodies, very similar levels of receptor dimerization were obtained (Movie S3 and Figure 4D). A slightly increased dimerization level compared to EPO was observed, which may be due to the symmetric binding affinities of diabodies to both EpoR subunits compared to the asymmetric receptor dimer assembly observed for EPO. Importantly, the diffusion properties of receptor dimers assembled by the diabodies were comparable to EPO, as shown for DA5 in Figure 4E, confirming a comparable mode of receptor dimerization by diabodies compared to EPO. Moreover, 1:1 receptor dimers recruited by the diabodies is observed by single-step photobleaching (Figure 4F). Thus, although we do not rule out any role of EpoR pre-association in the observed signaling effects, our microscopy data indicate that the diabodies are not simply clustering quiescent EpoR dimers into higher-order assemblies.

### EpoR Diabodies Inhibit Erythroid Colony Formation in JAK2V617F-Positive Patients

Several mutations in JAKs are known to cause immune disorders and cancer by rendering activation ligand independent (Gäbler et al., 2013; James et al., 2005). We asked whether the large EpoR distances and different binding geometries induced by the diabodies could modulate the activity of these kinase mutants in an extracellular ligand-dependent manner by separating the two JAKs at distances where they could not undergo trans-activation. The JAK2V617F mutant is the best-described example of an oncogenic JAK mutation, causing the development of hematological disorders such as polycythemia vera (PV) and other myeloproliferative (MPN) neoplasms (Baxter et al., 2005; James et al., 2005; Kralovics et al., 2005; Levine et al., 2005). At physiologic expression levels, JAK2V617F-positive cells require EpoR to proliferate in a ligand-independent manner (Lu et al., 2008). Stimulation of Ba/F3 cells expressing the murine EpoR and the JAK2V617F mutant with EPO or DA5 did not significantly affect the basal phosphorylation of STAT5, Akt, and Erk in these cells (Figures 5B and 5C). However, stimulation of these cells with DA10, DA307, and DA330 decreased the STAT5, Akt, and Erk phosphorylation in a time-dependent manner (Figures 5B and 5C). This decrease in signal activation induced by DA10, DA307, and DA330 was not the result of EpoR surface depletion. Only the full agonists, EPO and DA5, led to a significant decrease in the levels of EpoR on the surface (Figure 5D). The decrease in the JAK2V617F-induced basal signaling activation promoted by the diabodies was followed by a reduction in the proliferation rate of Ba/F3 cells expressing the mutated JAK2 (Figure 5E), suggesting that oncogenic JAK



**Figure 4. Diabodies Dimerize EpoR at the Surface of Living Cells**

(A) Cell-surface labeling of EpoR using dye-labeled anti-GFP nanobodies.

(B) Relative co-localization of RHO11EpoR and DY647EpoR in absence and presence of ligand. As a negative control, co-localization of maltose binding protein fused to an indifferent trans-membrane domain is shown. Data (mean  $\pm$  max/min) are shown.

(C) Trajectories (150 frames,  $\sim$ 4.8 s) of individual Rho11-labeled (red) and DY647-labeled EpoR (blue) and co-trajectories (magenta) for unstimulated cells, as well as after stimulation with EPO (5 nM) and DA5 (250 nM).

(D) Relative amount of co-trajectories for unstimulated EPOR and after stimulation with EPO and diabodies (DA5, DA330, and DA10). Data (mean  $\pm$  max/min) are shown.

(E) Diffusion properties of EpoR represented as trajectory step-length distribution (time lapse: 160 ms) for unstimulated cells and after dimerization with EPO or DA5. The curves correspond to fitted data from  $>10$  cells ( $\sim$ 1,500 trajectories each).

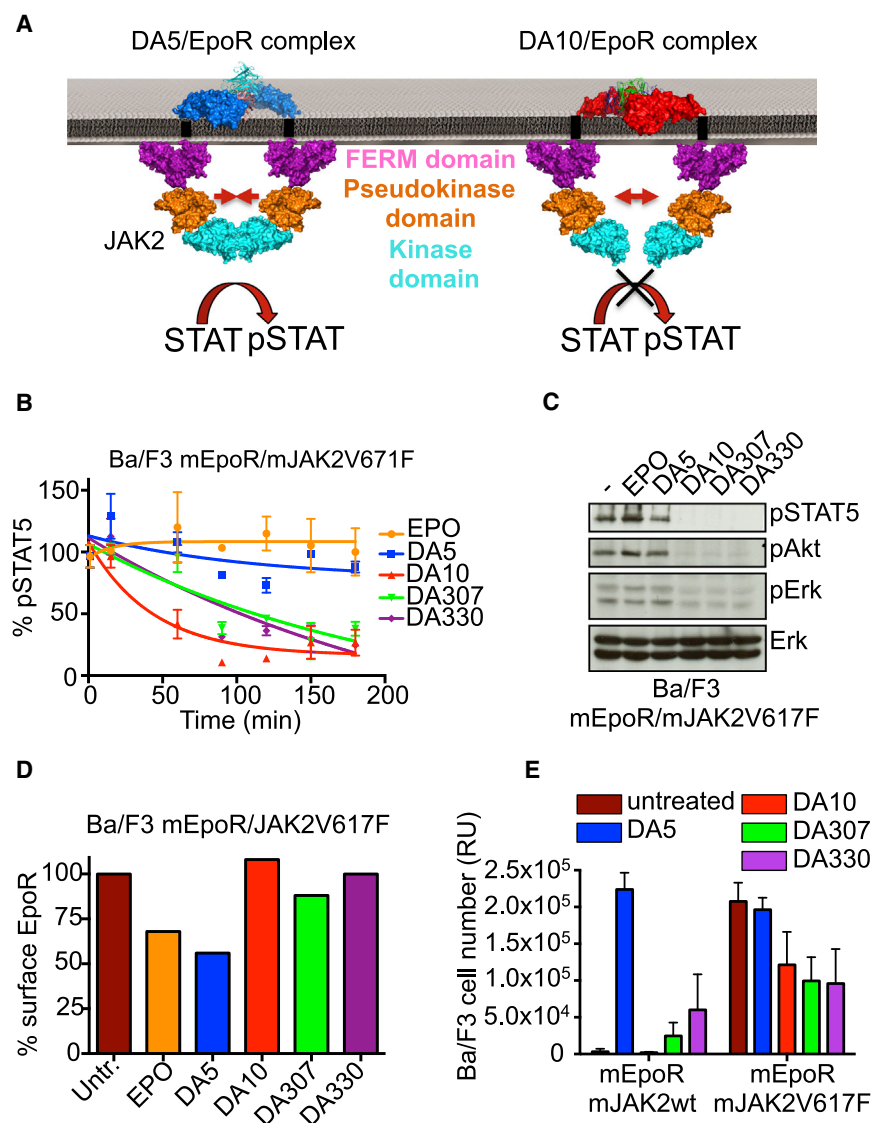
(F) Diabody-induced dimerization of EpoR demonstrated by dual-step bleaching analysis. Top: a pseudo-3D kymograph illustrating dual-color single-step bleaching for an individual DA5-induced EpoR co-trajectory. Bottom left: the corresponding pixel-intensity profiles are shown for both acquisition channels. Bottom right: the fraction of signals within co-trajectories that decay within a single step versus multiple steps. Comparison for complexes obtained with EPO (from 154 co-trajectories) and DA5 (from 186 co-trajectories).

mutant activities can be modulated in an extracellular ligand-dependent manner.

The Ba/F3 cells used here are a transformed cell line engineered to overexpress EpoR and JAK2V617F, which led to transformation and to autonomous growth, so we also performed erythroid colony formation assays in primary cells from human JAK2V617F-positive patients. CD34<sup>+</sup> hematopoietic stem cells and progenitors (HSPC) from heterozygous JAK2V617F-positive patients were isolated and stimulated with the indicated diabodies  $\pm$  EPO, and their ability to form erythroid colonies was assayed. In the absence of diabodies, JAK2V617F-positive CD34<sup>+</sup> cells gave rise to erythroid colonies, which were further increased in numbers in the presence of EPO in the media (Figure 6A). Stimulation with a non-specific negative control diabody did not significantly alter the number of erythroid (EpoR-dependent) or myeloid colonies (EpoR independent) (Figures 6A and 6B), ruling out possible toxic side effects induced by

diabodies. Stimulation of JAK2V617F-positive CD34<sup>+</sup> cells with the agonistic diabody DA5 led to a specific increase in the number of erythroid colonies (Figures 6A and 6C) without significantly altering the number of myeloid colonies (Figure 6B). On the other hand, stimulation with DA330 and DA10 led to a potent and specific decrease in the number of erythroid colonies (Figures 6A–6C). We note that DA330, which is a partial agonist of normal JAK2 signaling, limits but does not prevent signaling in JAK2V617F cells, giving the appearance of a structural “governor” controlling signaling output. All of the colonies analyzed in the study harbored the JAK2V617F mutation as determined by single-colony genotyping (Figure 6D). The diabody with the largest intersubunit distance, DA10, inhibited colony formation the strongest, comparably to the JAK1/2 inhibitor Ruxolitinib, which is approved and standard of care for JAK2V617F-positive MPN (Verstovsek et al., 2010) (Figure 6A). DA10 also decreased the number of erythroid colonies from homozygous JAK2V617F-positive patients (Dupont et al., 2007) (Figures 6E and 6F), suggesting that the binding topology imposed by this diabody dominates over the influence of the mutated JAK2 expressed in the cell. Overall, these results show that extracellular ligands





**Figure 5. DA10 and DA330 Inhibit JAK2V617F Constitutive Activity**

(A) Model depicting the mechanism by which the diabodies affect signaling activation potencies. The large dimer intersubunit distances exhibited by the diabodies may alter the position of JAK2 upon ligand binding, decreasing its ability to transactivate each other and start downstream signaling amplification.

(B) Kinetics of pSTAT5 in Ba/F3 cells expressing the JAK2V617F mutant after stimulation with EPO or the four diabodies. DA10, DA307, and DA330 induce a decrease on the basal pSTAT5 levels in a time-dependent manner. Data (mean  $\pm$  SD) are from two independent experiments.

(C) pSTAT5, pErk, and pAkt levels induced by 1  $\mu$ M of the four diabodies in Ba/F3 cells expressing the JAK2V617F mutant after 3 hr of stimulation.

(D) EpoR surface levels after 1 hr stimulation with EPO or the four diabodies.

(E) Proliferation of Ba/F3 cells expressing JAK2 WT or JAK2V617F in response to 1  $\mu$ M of each of the four diabodies after 5 days of stimulation. Data (mean  $\pm$  SD) are from three independent experiments.

that enforce large receptor dimer separation and different binding geometries can counteract intracellular oncogenic ligand-independent receptor activation, presumably by exceeding the accessible distance that the JAK2 kinase domain can extend to transphosphorylate the opposing JAK2 and receptor (Figure 5A).

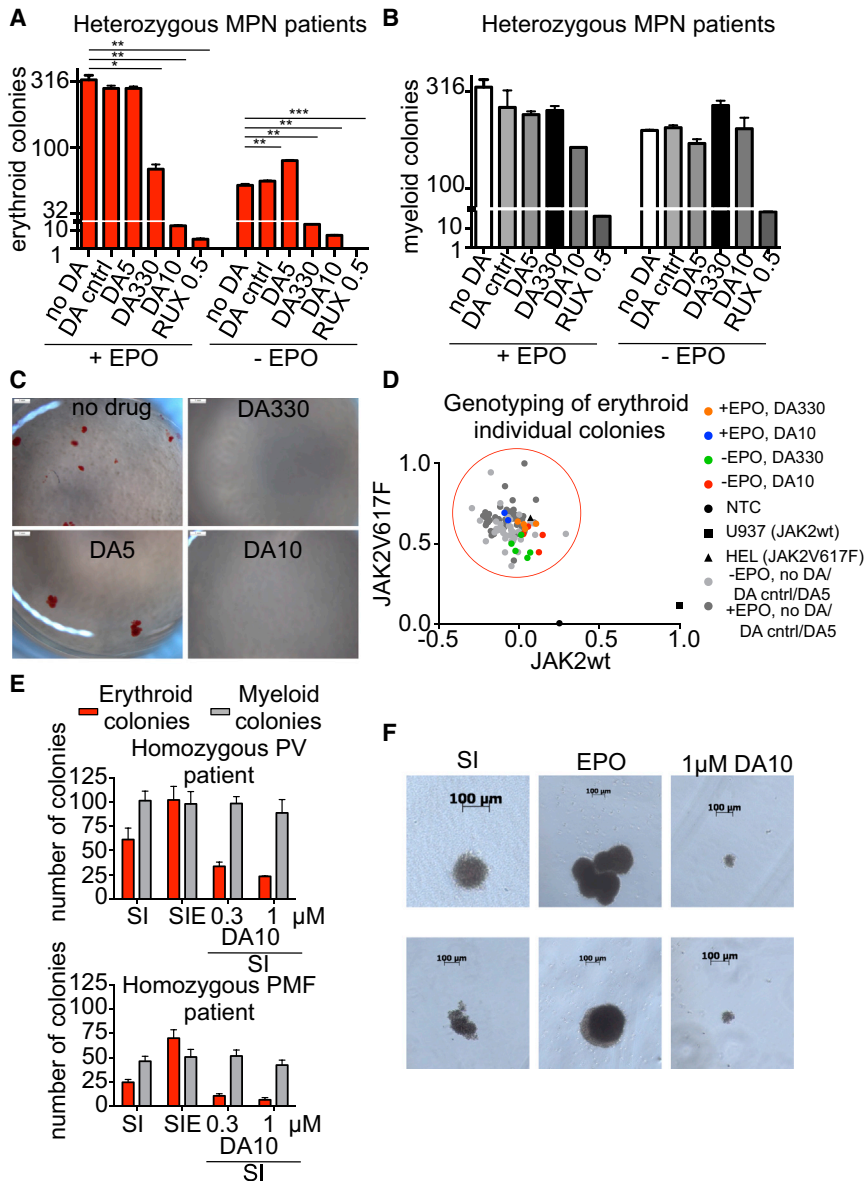
## DISCUSSION

Single-pass type I and type II transmembrane receptors that contain ligand-binding ECDs constitute a major percentage of all signaling receptors in the mammalian genome and include cytokine (JAK/STAT) receptors (Spangler et al., 2014), tyrosine kinase (RTK) receptors (e.g., EGF-R, Insulin-R, etc.) (Lemmon and Schlessinger, 2010), and many others. In most cases, these receptors signal in response to ligand engagement as homo- or heterodimeric units (Klemm et al., 1998; Stroud and Wells, 2004).

For this class of receptors, ligand binding ECDs are structurally autonomous and are separated from the intracellular signaling modules (e.g., Kinase domains) through juxtamembrane linkers and a TM helix. Thus, the intracellular domains (ICDs) presumably sense ligand binding through spatial perturbations of receptor orientation and proximity that are relayed as conformational changes through the membrane (Ottemann et al., 1999). However, it has been unclear to what extent extracellular ligands can influence signaling through dimeric receptors by enforcing ECD orientational differences. In contrast, for GPCRs, although the role of dimerization remains to be determined, it is well established that ligand binding within the TM helices induces conformational changes within the plane of the membrane. Even minor structural differences in the relative orientations of GPCR TM helices induced by ligands are conveyed as differential signaling (e.g., biased signaling, inverse and partial agonism) (Venkatakrishnan et al., 2013). This property of GPCRs has been exploited by the pharmaceutical industry for small-molecule drug development. Here, we asked whether ligand-induced orientational (i.e., “shape”) changes of receptor dimer geometry could serve a conceptually and functionally analogous role to the diverse types of conformational changes induced by GPCR ligands that result in differential signaling.

Although there exists a vast literature showing that dimeric receptor signaling strength is determined by extracellular parameters such as ligand affinity and complex half-life on the cell surface (Harwerth et al., 1992; Riese, 2011), the role of





**Figure 6. DA10 and DA330 Inhibit Erythroid Colony Formation in JAK2V617F-Positive Patient Samples**

(A) Number of erythroid BFU-E (EpoR-dependent) colonies in heterozygous JAK2V617F positive myeloproliferative neoplasm patient samples after stimulation with the indicated ligands. Data (mean  $\pm$  SD) are from three different donors.

(B) Number of myeloid colonies in heterozygous JAK2V617F-positive myeloproliferative neoplasm patient samples after stimulation with the indicated ligands.

(C) Overview pictures highlight EPO-independent BFU-E colonies (no drug and DA5), which are significantly diminished with DA330 and DA10 treatment. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ ; paired Student's *t* test was used to determine significant changes.

(D) The genotype of 109 erythroid colonies derived from sorted CD34+ cells derived from PMF cases was determined by multiplexed custom TaqMan SNP assay for JAK2V617F and JAK2 wild-type. Each colony is represented by a single dot in the graph and colored according to different treatment regimens. Gray dots represent colonies derived from conditions without treatment or treatment with an agonist (dark gray with EPO, light gray without EPO), orange and green dots represent few residual colonies treated with DA330, and blue and red dots very rare residual colonies treated with DA10.

(E) Number of erythroid colonies (burst-forming units-erythroid [BFU] or endogenous erythroid colonies [EEC]) and myeloid colonies (EpoR-independent) in a polycythemia vera (PV) (top) and primary myelofibrosis (PMF) patient (bottom panel) homozygous for JAK2V617F. SI: SCF + IL-3; SIE: SCF + IL-3 + EPO. Data (mean  $\pm$  SD) are from three different donors.

(F) Morphology of EEC colonies after treatment with the indicated conditions is shown.

orientation-specific effects has remained speculative (Ballinger and Wells, 1998; Syed et al., 1998; Wells and de Vos, 1993). Studies using mutated, chimeric, or genetically modified receptors have pointed to the importance of the extracellular domain structure in mediating signaling output (Barclay et al., 2010; Liu et al., 2009; Millot et al., 2004; Rowlinson et al., 2008; Seubert et al., 2003; Staerk et al., 2011). Nevertheless, for this parameter to be exploited in a manner that could be useful therapeutically, surrogate ligands with the capacity to induce alternative signaling outputs through naturally, non-mutated receptors on human cells are required. We used diabodies because they would presumably induce large-scale alterations in dimer geometry and have been previously shown to have the capacity to act as agonists of c-MPL (Nakano et al., 2009). Although antibodies have been shown to elicit diverse functional and signaling out-

puts through cytokine receptors (Zhang et al., 2012a; Kai et al., 2008), they are elusive structural targets due to their segmental flexibility. Diabodies have more constrained structures than antibodies (Perisic et al., 1994), which allowed us to capture the receptor-diabody signaling complexes crystallographically.

Our results indicate that cytokine receptor dimer architectural and spacing constraints compatible with signaling are liberal but there exist limits at which signaling is impacted. This is consistent with the diverse range of dimeric ligand-receptor geometries seen in agonistic cytokine-receptor complex structures (Span- gler et al., 2014; Wang et al., 2009). Consequently, we find that large-scale re-orientations of receptor dimer topology are required to qualitatively and quantitatively modulate signaling output. We propose that this strategy is potentially applicable to other dimeric receptor systems, such as RTKs, where the role of ligand is to bind to the ECDs, dimerize, and/or re-orient receptors.

The broader implications of our results are that signaling patterns delivered by endogenous ligands only constitute one of many possible signaling patterns that can be elicited through a dimeric receptor system. By using surrogate or engineered ligands to re-orient receptor dimer topology, a given dimeric receptor can be induced to deliver a wide range of signals of different amplitudes and pathway specificities. Cytokine receptor dimers have the potential to be modulated as rheostats to control signaling output, similar to partial and biased GPCR agonists. Given that many endogenous cytokines and growth factors have adverse effects as therapeutic agonists, our results portend the possibility of dimer re-orientation as a strategy to “tune” signaling output to minimize toxicity, maximize efficacy, or elicit specific functional outcomes.

The precise molecular mechanisms through which the diabodies described here alter intracellular signaling by remodeling dimer geometry remain unclear, but the signal tuning effects are clearly the result of extracellular receptor dimer proximity (distance) and geometry (orientation) effects. Our single-molecule fluorescence tracking shows that the assembled signaling complexes are not due to higher-order assemblies that could have resulted from diabody-induced clustering of preformed EpoR receptor dimers (Constantinescu et al., 2001). Even if receptor clustering were occurring to some degree, which we do not rule out, the diabodies still exert a powerful modulatory effect on signaling through repositioning receptor topology whether or not these are monovalent or polyvalent cell-surface complexes. This strategy does not rely on a particular valency of the signaling complexes. For example, our results can be reconciled with a recent mechanistic study of cytokine receptor activation (Brooks et al., 2014). Growth hormone was shown to activate its receptor (GH-R) by rotating the ECD subunits of a pre-associated but inactive GH-R dimer, resulting in separation of the Box 1 receptor ICD motifs and removal of the JAK2 pseudokinase inhibitory domain, which collectively result in productive JAK2 kinase domain positioning for receptor activation. Diabodies could presumably disrupt a quiescent cytokine receptor dimer to form an activated dimer topology through a related “separation” mechanism that relieves JAK2 inhibition. For the agonist DA5, the outcome of this separation would be placement of the JAK kinase domains into productive apposition but one that is topologically distinct from that induced by the natural cytokine. In the case of DA10, the kinase domains of JAK2 are separated such that they are not in proper position to trans-phosphorylate. We contend that such a JAK activation mechanism could still be operative in the context of non-native dimer architectures.

The surprisingly large EpoR dimer separation distances imposed by the agonistic diabodies may be rationalized by the fact that the intracellular, receptor-associated JAKs are long molecules that exist as a dynamic ensemble of extended and compact conformations, which could span >100Å distances between receptors in a dimer (Lupardus et al., 2011). Given that the kinase domain of JAK resides at its C terminus, which is most distal to the receptor bound by the JAK FERM domain, it is likely sensitive to positioning relative to its substrates that it trans-phosphorylates. Changes in the relative positioning of the kinase domain to its substrates could influence the efficiency and patterns of phosphorylation through steric effects imposed by

extracellular dimer geometry. By manipulating the dimer geometry, as seen with the non-signaling diabody DA10, such an approach can achieve complete shutoff of constitutively active signaling pathways (JAK2V617F) from the outside of the cell. This is conceptually distinct from Ankyrin repeat antagonists to ErbB2 that were shown to prevent activation of wild-type ErbB2 by distorting the receptors such that they cannot form signaling-competent dimers (Jost et al., 2013). Here, the role of DA10 is to dimerize EpoR yet terminate ligand-independent signaling, possibly through enforcing a large dimer separation distance. This strategy is applicable to diseases mediated by mutated, constitutively active receptors (Bivona et al., 2011; Pikman et al., 2006; Rebouissou et al., 2009; Zenatti et al., 2011) and could offer the advantage of specificity and reduced toxicity versus broadly neutralizing kinase inhibitors.

Diabodies are a convenient surrogate ligand because they can be created from existing monoclonal antibody sequences, which exist to most human cell-surface receptors. However, dimer re-orientation could be achieved by many different types of engineered scaffolds. A range of altered dimerization geometries could be screened with different dimerizing scaffolds for those that induced a particular signaling profile or functional property. In principle, targeting receptor ECD dimer orientation as a new structure-activity parameter for drug discovery for many type I or type II cell-surface receptors is feasible.

## EXPERIMENTAL PROCEDURES

Further details for production, characterization, and crystallization of diabodies; signaling and functional characterization; in vivo imaging of surface DA-EpoR complex formation; and isolation and treatment of JAK2V617F-positive human samples can be found online in the [Extended Experimental Procedures](#).

### Structure Determination and Refinement

All crystallographic data were collected at the Stanford Synchrotron Radiation Lightsources (Stanford) beamlines 12-2. Data were indexed, integrated and scaled using XDS or HKL2000 program suits (Kabsch, 2010; Otwinowski et al., 1997). The three DA-EpoR crystal structures were solved by molecular replacement with the program PHASER (McCoy, 2007) and refined with PHENIX and COOT.

### Primarity Bio Pathway Phenotyping

UT-7-EpoR cells were starved overnight; stimulated with saturated concentrations of EPO and the indicated diabodies for 15, 60, and 120 min; and fixed with 1% PFA for 10 min at room temperature. The fixed cells were prepared for antibody staining according to standard protocols (Krutzik and Nolan, 2003). Briefly, the fixed cells were permeabilized in 90% methanol for 15 min. The cells were stained with a panel of antibodies specific to the markers indicated (Primarity Bio Pathway Phenotyping service and [Table S1](#)) and analyzed on an LSRII flow cytometer (Becton Dickinson). The Log2 Ratio of the median fluorescence intensities (MFI) of the stimulated samples divided by the unstimulated control samples were calculated as a measure of response.

### Single-Molecule Tracking, Co-localization, and Co-tracking Analyses

Single-molecule localization and single-molecule tracking were carried out using the multiple-target tracing (MTT) algorithm (Sergé et al., 2008) as described previously (You et al., 2010). Step-length histograms were obtained from single-molecule trajectories and fitted by a two fraction mixture model of Brownian diffusion. Average diffusion constants were determined from the slope (2–10 steps) of the mean square displacement versus time lapse

diagrams. Immobile molecules were identified by the density-based spatial clustering of applications with noise (DBSCAN) (Sander et al., 1998) algorithm as described recently (Roder et al., 2014). For comparing diffusion properties and for co-tracking analysis, immobile particles were excluded from the data set. Individual molecules detected in the both spectral channels were regarded as co-localized if a particle was detected in both channels of a single frame within a distance threshold of 100 nm radius.

### HSC and Progenitor-Derived Colony Genotyping Assay

CD34+ cells were sorted from human JAK2V617F homo- and heterozygous myeloproliferative samples. CD34+ cells were plated in methylcellulose with and without erythropoietin (MethoCult H4434 and H4535; STEMCELL Technologies). Colony formation was assessed after 14 days in culture by microscopy and scored on the basis of morphology. JAK2V617F and JAK2 WT TaqMan SNP Genotyping Assay (Applied Biosystems) was designed as published recently (Levine et al., 2006), and details are available upon request. The genotype of each colony was determined by Custom TaqMan SNP Genotyping Assay (Applied Biosystems) according to the manufacturer's specification.

### ACCESSION NUMBERS

The Protein Data Bank (PDB) accession numbers for the three DA/EpoR complex structures reported in this paper are 4Y5V (DA5-EpoR), 4Y5X (DA10-EpoR), and 4Y5Y (DA330-EpoR). RNA-seq data can be accessed via National Center for Biotechnology Information (NCBI) BioProject under the accession number PRJNA275804.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, two tables, and five movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.011>.

### AUTHOR CONTRIBUTIONS

I.M. and K.C.G. conceived the project and wrote the manuscript. I.M., T.S.W., H.F., V.G., and P.K. functionally characterized the diabodies. S.W., C.P.R., and J.P. designed and performed single-particle microscopy experiments. V.G., S.N.C., I.P., G.W., W.-J.H., R.M., and I.L.W. designed and performed experiments in human JAK2V617F-positive patients. G.W. and R.S. performed RNA-seq libraries and analysis. I.M. and F.G. expressed recombinant proteins and performed crystallization experiments of DA-EpoR complexes. S.D. and P.M. designed and performed molecular dynamics experiments.

### ACKNOWLEDGMENTS

We thank Kevin Jude for technical assistance; the Division of Hematology Tissue Bank and the patients for donating their samples; Dr. William Vainchenker, Institut Gustave Roussy for providing patient JAK2V617F homozygous CD34+ cells; and Dr. Christian Pecquet for support with cell lines. This work was supported by the BELSPO and de Duve Institute Delori postdoctoral fellowships (V.G.), T32CA009287-35 (Training Program in Investigative Oncology) and TG2-01159 (Stanford CIRM Training Program) (to W.I.H.), NIH U01HL099999 (to G.W., R.S., and I.L.W.), FRS-FNRS, Salus Sanguinis Foundation, Fondation contre le cancer (Belgium), Programs IAP-BeMGI and ARC (S.N.C.), the Ludwig Institute (K.C.G.), NIH-RO1-AI51321 (K.C.G.), and the Howard Hughes Medical Institute (K.C.G.).

Received: December 4, 2014

Revised: January 22, 2015

Accepted: February 3, 2015

Published: February 26, 2015

### REFERENCES

Ballinger, M.D., and Wells, J.A. (1998). Will any dimer do? *Nat. Struct. Biol.* 5, 938–940.

Barclay, J.L., Kerr, L.M., Arthur, L., Rowland, J.E., Nelson, C.N., Ishikawa, M., d'Aniello, E.M., White, M., Noakes, P.G., and Waters, M.J. (2010). In vivo targeting of the growth hormone receptor (GHR) Box1 sequence demonstrates that the GHR does not signal exclusively through JAK2. *Mol. Endocrinol.* 24, 204–217.

Baxter, E.J., Scott, L.M., Campbell, P.J., East, C., Fourouclas, N., Swanton, S., Vassiliou, G.S., Bench, A.J., Boyd, E.M., Curtin, N., et al.; Cancer Genome Project (2005). Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 365, 1054–1061.

Bazan, J.F. (1990). Haemopoietic receptors and helical cytokines. *Immunol. Today* 11, 350–354.

Bivona, T.G., Hieronymus, H., Parker, J., Chang, K., Taron, M., Rosell, R., Moonsamy, P., Dahlman, K., Miller, V.A., Costa, C., et al. (2011). FAS and NF- $\kappa$ B signalling modulate dependence of lung cancers on mutant EGFR. *Nature* 471, 523–526.

Boersma, Y.L., Chao, G., Steiner, D., Wittrup, K.D., and Plückthun, A. (2011). Bispecific designed ankyrin repeat proteins (DARPs) targeting epidermal growth factor receptor inhibit A431 cell proliferation and receptor recycling. *J. Biol. Chem.* 286, 41273–41285.

Boulanger, M.J., Chow, D.C., Brevnova, E.E., and Garcia, K.C. (2003). Hexameric structure and assembly of the interleukin-6/IL-6  $\alpha$ -receptor/gp130 complex. *Science* 300, 2101–2104.

Brooks, A.J., Dai, W., O'Mara, M.L., Abankwa, D., Chhabra, Y., Pelekanos, R.A., Gardon, O., Tunny, K.A., Blucher, K.M., Morton, C.J., et al. (2014). Mechanism of activation of protein kinase JAK2 by the growth hormone receptor. *Science* 344, 1249783.

Constantinescu, S.N., Ghaffari, S., and Lodish, H.F. (1999). The Erythropoietin Receptor: Structure, Activation and Intracellular Signal Transduction. *Trends Endocrinol. Metab.* 10, 18–23.

Constantinescu, S.N., Huang, L.J., Nam, H., and Lodish, H.F. (2001). The erythropoietin receptor cytosolic juxtamembrane domain contains an essential, precisely oriented, hydrophobic motif. *Mol. Cell* 7, 377–385.

de Vos, A.M., Ultsch, M., and Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. *Science* 255, 306–312.

Decker, T., and Kovarik, P. (2000). Serine phosphorylation of STATs. *Oncogene* 19, 2628–2637.

Drake, M.T., Violin, J.D., Whalen, E.J., Wisler, J.W., Shenoy, S.K., and Lefkowitz, R.J. (2008).  $\beta$ -arrestin-biased agonism at the  $\beta$ 2-adrenergic receptor. *J. Biol. Chem.* 283, 5669–5676.

Dupont, S., Massé, A., James, C., Teyssandier, I., Lécluse, Y., Larbret, F., Ugo, V., Saulnier, P., Koscielny, S., Le Couédic, J.P., et al. (2007). The JAK2 617V>F mutation triggers erythropoietin hypersensitivity and terminal erythroid amplification in primary cells from patients with polycythemia vera. *Blood* 110, 1013–1021.

Gäbler, K., Behrmann, I., and Haan, C. (2013). JAK2 mutants (e.g., JAK2V617F) and their importance as drug targets in myeloproliferative neoplasms. *JAK-STAT* 2, e25025.

Gent, J., van Kerkhof, P., Roza, M., Bu, G., and Strous, G.J. (2002). Ligand-independent growth hormone receptor dimerization occurs in the endoplasmic reticulum and is required for ubiquitin system-dependent endocytosis. *Proc. Natl. Acad. Sci. USA* 99, 9858–9863.

Hansen, G., Hercus, T.R., McClure, B.J., Stomski, F.C., Dottore, M., Powell, J., Ramshaw, H., Woodcock, J.M., Xu, Y., Guthridge, M., et al. (2008). The structure of the GM-CSF receptor complex reveals a distinct mode of cytokine receptor activation. *Cell* 134, 496–507.

Harwerth, I.M., Wels, W., Marte, B.M., and Hynes, N.E. (1992). Monoclonal antibodies against the extracellular domain of the erbB-2 receptor function as partial ligand agonists. *J. Biol. Chem.* 267, 15160–15167.

Heller, N.M., Qi, X., Gesbert, F., and Keegan, A.D. (2012). The extracellular and transmembrane domains of the  $\gamma$ C and interleukin (IL)-13 receptor  $\alpha$ 1 chains, not their cytoplasmic domains, dictate the nature of signaling responses to IL-4 and IL-13. *J. Biol. Chem.* 287, 31948–31961.

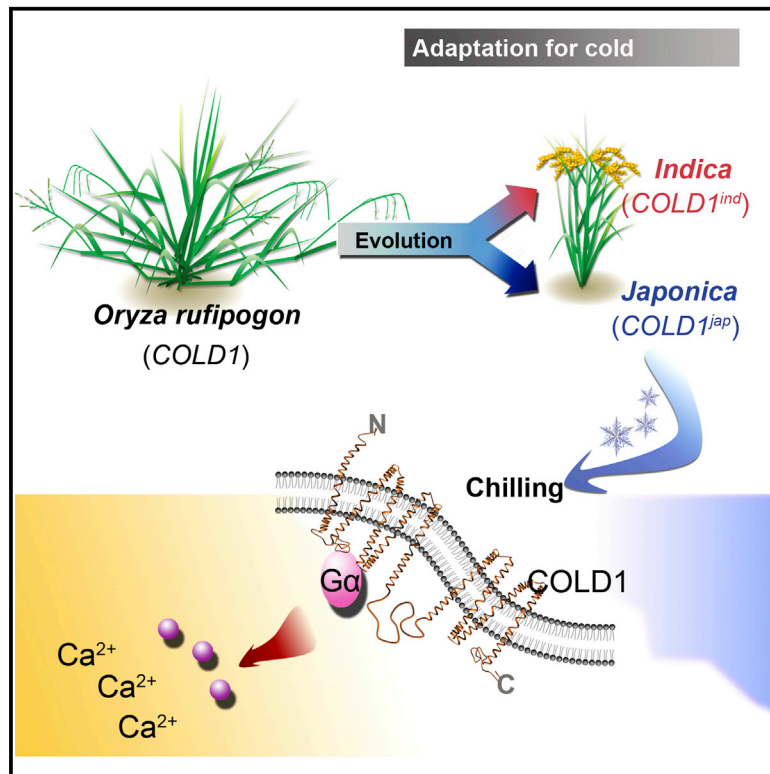
- Ihle, J.N., Witthuhn, B.A., Quelle, F.W., Yamamoto, K., and Silvennoinen, O. (1995). Signaling through the hematopoietic cytokine receptors. *Annu. Rev. Immunol.* **13**, 369–398.
- James, C., Ugo, V., Casadevall, N., Constantinescu, S.N., and Vainchenker, W. (2005). A JAK2 mutation in myeloproliferative disorders: pathogenesis and therapeutic and scientific prospects. *Trends Mol. Med.* **11**, 546–554.
- Jost, C., Schilling, J., Tamaskovic, R., Schwill, M., Honegger, A., and Plückthun, A. (2013). Structural basis for eliciting a cytotoxic effect in HER2-overexpressing cancer cells via binding to the extracellular domain of HER2. *Structure* **21**, 1979–1991.
- Kabsch, W. (2010). Xds. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132.
- Kai, M., Motoki, K., Yoshida, H., Emuta, C., Chisaka, Y., Tsuruhata, K., Endo, C., Muto, M., Shimabe, M., Nishiyama, U., et al. (2008). Switching constant domains enhances agonist activities of antibodies to a thrombopoietin receptor. *Nat. Biotechnol.* **26**, 209–211.
- Kavran, J.M., McCabe, J.M., Byrne, P.O., Connacher, M.K., Wang, Z., Ramek, A., Sarabipour, S., Shan, Y., Shaw, D.E., Hristova, K., et al. (2014). How IGF-1 activates its receptor. *eLife* **3**, e03772.
- Klemm, J.D., Schreiber, S.L., and Crabtree, G.R. (1998). Dimerization as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.* **16**, 569–592.
- Kralovics, R., Passamonti, F., Buser, A.S., Teo, S.S., Tiedt, R., Passweg, J.R., Tichelli, A., Cazzola, M., and Skoda, R.C. (2005). A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* **352**, 1779–1790.
- Krutzik, P.O., and Nolan, G.P. (2003). Intracellular phospho-protein staining techniques for flow cytometry: monitoring single cell signaling events. *Cytometry A* **55**, 61–70.
- LaPorte, S.L., Juo, Z.S., Vaclavikova, J., Colf, L.A., Qi, X., Heller, N.M., Keegan, A.D., and Garcia, K.C. (2008). Molecular and structural basis of cytokine receptor pleiotropy in the interleukin-4/13 system. *Cell* **132**, 259–272.
- Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134.
- Levine, R.L., Wadleigh, M., Cools, J., Ebert, B.L., Wernig, G., Huntly, B.J., Boggon, T.J., Wlodarska, I., Clark, J.J., Moore, S., et al. (2005). Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* **7**, 387–397.
- Levine, R.L., Belisle, C., Wadleigh, M., Zahrieh, D., Lee, S., Chagnon, P., Gilliland, D.G., and Busque, L. (2006). X-inactivation-based clonality analysis and quantitative JAK2V617F assessment reveal a strong association between clonality and JAK2V617F in PV but not ET/MMM, and identifies a subset of JAK2V617F-negative ET and MMM patients with clonal hematopoiesis. *Blood* **107**, 4139–4141.
- Li, W., Lan, H., Liu, H., Fu, Z., Yang, Y., Han, W., Guo, F., Liu, Y., Zhang, H., Liu, J., and Zheng, X. (2013). The activation and differential signalling of the growth hormone receptor induced by pGH or anti-idiotypic monoclonal antibodies in primary rat hepatocytes. *Mol. Cell. Endocrinol.* **376**, 51–59.
- Liao, W., Lin, J.X., and Leonard, W.J. (2011). IL-2 family cytokines: new insights into the complex roles of IL-2 as a broad regulator of T helper cell differentiation. *Curr. Opin. Immunol.* **23**, 598–604.
- Lim, A.C., Ketchum, R.R., Borges, L., Carabeo, T., Carter, J., Hoover, J.E., Hu, Z., Wittekind, M., Zhou, H., and Mehlin, C. (2010). A diversity of antibody epitopes can induce signaling through the erythropoietin receptor. *Biochemistry* **49**, 3797–3804.
- Liu, W., Kawahara, M., Ueda, H., and Nagamune, T. (2009). The influence of domain structures on the signal transduction of chimeric receptors derived from the erythropoietin receptor. *J. Biochem.* **145**, 575–584.
- Livnah, O., Stura, E.A., Johnson, D.L., Middleton, S.A., Mulcahy, L.S., Wrighton, N.C., Dower, W.J., Jolliffe, L.K., and Wilson, I.A. (1996). Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science* **273**, 464–471.
- Livnah, O., Johnson, D.L., Stura, E.A., Farrell, F.X., Barbone, F.P., You, Y., Liu, K.D., Goldsmith, M.A., He, W., Krause, C.D., et al. (1998). An antagonist peptide-EPO receptor complex suggests that receptor dimerization is not sufficient for activation. *Nat. Struct. Biol.* **5**, 993–1004.
- Livnah, O., Stura, E.A., Middleton, S.A., Johnson, D.L., Jolliffe, L.K., and Wilson, I.A. (1999). Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **283**, 987–990.
- Lu, X., Huang, L.J., and Lodish, H.F. (2008). Dimerization by a cytokine receptor is necessary for constitutive activation of JAK2V617F. *J. Biol. Chem.* **283**, 5258–5266.
- Lupardus, P.J., Skiniotis, G., Rice, A.J., Thomas, C., Fischer, S., Walz, T., and Garcia, K.C. (2011). Structural snapshots of full-length Jak1, a transmembrane gp130/IL-6/IL-6R $\alpha$  cytokine receptor complex, and the receptor-Jak1 holo-complex. *Structure* **19**, 45–55.
- McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41.
- Millot, G.A., Vainchenker, W., Duménil, D., and Svinarchuk, F. (2004). Differential signalling of NH2-terminal flag-labelled thrombopoietin receptor activated by TPO or anti-FLAG antibodies. *Cell. Signal.* **16**, 355–363.
- Müller-Newen, G., Küster, A., Wijdenes, J., Schaper, F., and Heinrich, P.C. (2000). Studies on the interleukin-6-type cytokine signal transducer gp130 reveal a novel mechanism of receptor activation by monoclonal antibodies. *J. Biol. Chem.* **275**, 4579–4586.
- Nakano, K., Kojima, T., Kasutani, K., Senoh, C., Natori, O., Ishii, S., Tsunoda, H., and Hattori, K. (2009). Effective screening method of agonistic diabetides based on autocrine growth. *J. Immunol. Methods* **347**, 31–35.
- O'Shea, J.J., and Paul, W.E. (2010). Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. *Science* **327**, 1098–1102.
- Ohashi, H., Maruyama, K., Liu, Y.C., and Yoshimura, A. (1994). Ligand-induced activation of chimeric receptors between the erythropoietin receptor and receptor tyrosine kinases. *Proc. Natl. Acad. Sci. USA* **91**, 158–162.
- Ottemann, K.M., Xiao, W., Shin, Y.K., and Koshland, D.E., Jr. (1999). A piston model for transmembrane signaling of the aspartate receptor. *Science* **285**, 1751–1754.
- Otwinowski, Z., Minor, W., and Carter, C.W., Jr. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (New York: Academic Press), pp. 307–326.
- Pattyn, E., Van Ostade, X., Schauvliege, L., Verhee, A., Kalai, M., Vandekerckhove, J., and Tavernier, J. (1999). Dimerization of the interferon type I receptor IFNAR2-2 is sufficient for induction of interferon effector genes but not for full antiviral activity. *J. Biol. Chem.* **274**, 34838–34845.
- Perisic, O., Webb, P.A., Holliger, P., Winter, G., and Williams, R.L. (1994). Crystal structure of a diabody, a bivalent antibody fragment. *Structure* **2**, 1217–1226.
- Pikman, Y., Lee, B.H., Mercher, T., McDowell, E., Ebert, B.L., Gozo, M., Cuker, A., Wernig, G., Moore, S., Galinsky, I., et al. (2006). MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med.* **3**, e270.
- Rebouissou, S., Amessou, M., Couchy, G., Poussin, K., Imbeaud, S., Pilati, C., Izard, T., Balabaud, C., Bioulac-Sage, P., and Zucman-Rossi, J. (2009). Frequent in-frame somatic deletions activate gp130 in inflammatory hepatocellular tumours. *Nature* **457**, 200–204.
- Riese, D.J., 2nd. (2011). Ligand-based receptor tyrosine kinase partial agonists: New paradigm for cancer drug discovery? *Expert Opin. Drug Discov.* **6**, 185–193.
- Ring, A.M., Lin, J., Feng, D., Mitra, S., Rickert, M., Bowman, G.R., Pande, V.S., Li, P., Moraga, I., Spolski, R., et al. (2012). Mechanistic and structural insight into the functional dichotomy between IL-2 and IL-15. *Nat. Immunol.* **13**, 1187–1195.
- Roder, F., Wilmes, S., Richter, C.P., and Piehler, J. (2014). Rapid transfer of transmembrane proteins for single molecule dimerization assays in polymer-supported membranes. *ACS Chem. Biol.* **9**, 2479–2484.
- Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M.C., and Leonhardt, H. (2008). A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol. Cell. Proteomics* **7**, 282–289.
- Rowlinson, S.W., Yoshizato, H., Barclay, J.L., Brooks, A.J., Behncken, S.N., Kerr, L.M., Millard, K., Palethorpe, K., Nielsen, K., Clyde-Smith, J., et al.



- (2008). An agonist-induced conformational change in the growth hormone receptor determines the choice of signalling pathway. *Nat. Cell Biol.* 10, 740–747.
- Sander, J., Ester, M., Kriegel, H.P., and Xu, X.W. (1998). Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Min. Knowl. Discov.* 2, 169–194.
- Seita, J., Sahoo, D., Rossi, D.J., Bhattacharya, D., Serwold, T., Inlay, M.A., Ehrlich, L.I., Fathman, J.W., Dill, D.L., and Weissman, I.L. (2012). Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS ONE* 7, e40321.
- Sergé, A., Bertaux, N., Rigneault, H., and Marguet, D. (2008). Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* 5, 687–694.
- Seubert, N., Royer, Y., Staerk, J., Kubatzky, K.F., Moucadel, V., Krishnakumar, S., Smith, S.O., and Constantinescu, S.N. (2003). Active and inactive orientations of the transmembrane and cytosolic domains of the erythropoietin receptor dimer. *Mol. Cell* 12, 1239–1250.
- Socolovsky, M., Fallon, A.E., and Lodish, H.F. (1998). The prolactin receptor rescues EpoR<sup>-/-</sup> erythroid progenitors and replaces EpoR in a synergistic interaction with c-kit. *Blood* 92, 1491–1496.
- Spangler, J.B., Moraga, I., Mendoza, J.L., and Garcia, K.C. (2014). Insights into cytokine-receptor interactions from cytokine engineering. *Annu. Rev. Immunol.* Published online December 10, 2014. <http://dx.doi.org/10.1146/annurev-immunol-032713-120211>.
- Staerk, J., Defour, J.P., Pecquet, C., Leroy, E., Antoine-Poirel, H., Brett, I., Itaya, M., Smith, S.O., Vainchenker, W., and Constantinescu, S.N. (2011). Orientation-specific signalling by thrombopoietin receptor dimers. *EMBO J.* 30, 4398–4413.
- Stroud, R.M., and Wells, J.A. (2004). Mechanistic diversity of cytokine receptor signaling across cell membranes. *Sci. STKE* 2004, re7.
- Syed, R.S., Reid, S.W., Li, C., Cheetham, J.C., Aoki, K.H., Liu, B., Zhan, H., Osslund, T.D., Chirino, A.J., Zhang, J., et al. (1998). Efficiency of signalling through cytokine receptors depends critically on receptor orientation. *Nature* 395, 511–516.
- Thomas, C., Moraga, I., Levin, D., Krutzik, P.O., Podoplelova, Y., Trejo, A., Lee, C., Yarden, G., Vleck, S.E., Glenn, J.S., et al. (2011). Structural linkage between ligand discrimination and receptor activation by type I interferons. *Cell* 146, 621–632.
- Ullrich, A., and Schlessinger, J. (1990). Signal transduction by receptors with tyrosine kinase activity. *Cell* 61, 203–212.
- Venkatakrishnan, A.J., Deupi, X., Lebon, G., Tate, C.G., Schertler, G.F., and Babu, M.M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature* 494, 185–194.
- Verstovsek, S., Kantarjian, H., Mesa, R.A., Pardanani, A.D., Cortes-Franco, J., Thomas, D.A., Estrov, Z., Fridman, J.S., Bradley, E.C., Erickson-Viitanen, S., et al. (2010). Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N. Engl. J. Med.* 363, 1117–1127.
- Walter, M.R., Windsor, W.T., Nagabhushan, T.L., Lundell, D.J., Lunn, C.A., Zauodny, P.J., and Narula, S.K. (1995). Crystal structure of a complex between interferon-gamma and its soluble high-affinity receptor. *Nature* 376, 230–235.
- Wang, X., Rickert, M., and Garcia, K.C. (2005). Structure of the quaternary complex of interleukin-2 with its alpha, beta, and gamma receptors. *Science* 310, 1159–1163.
- Wang, X., Lupardus, P., Laporte, S.L., and Garcia, K.C. (2009). Structural biology of shared cytokine receptors. *Annu. Rev. Immunol.* 27, 29–60.
- Wehrman, T., He, X., Raab, B., Dukupatti, A., Blau, H., and Garcia, K.C. (2007). Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. *Neuron* 53, 25–38.
- Wells, J.A., and de Vos, A.M. (1993). Structure and function of human growth hormone: implications for the hematopoietins. *Annu. Rev. Biophys. Biomol. Struct.* 22, 329–351.
- You, C., Wilmes, S., Beutel, O., Löchte, S., Podoplelova, Y., Roder, F., Richter, C., Seine, T., Schaible, D., Uzé, G., et al. (2010). Self-controlled monofunctionalization of quantum dots for multiplexed protein tracking in live cells. *Angew. Chem. Int. Ed. Engl.* 49, 4108–4112.
- Zenatti, P.P., Ribeiro, D., Li, W., Zuurbier, L., Silva, M.C., Paganin, M., Tritapoe, J., Hixon, J.A., Silveira, A.B., Cardoso, B.A., et al. (2011). Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nat. Genet.* 43, 932–939.
- Zhang, H., Wilson, I.A., and Lerner, R.A. (2012a). Selection of antibodies that regulate phenotype from intracellular combinatorial antibody libraries. *Proc. Natl. Acad. Sci. USA* 109, 15728–15733.
- Zhang, H., Yea, K., Xie, J., Ruiz, D., Wilson, I.A., and Lerner, R.A. (2013). Selecting agonists from single cells infected with combinatorial antibody libraries. *Chem. Biol.* 20, 734–741.

# COLD1 Confers Chilling Tolerance in Rice

## Graphical Abstract



## Authors

Yun Ma, Xiaoyan Dai, ..., Song Ge, Kang Chong

## Correspondence

chongk@ibcas.ac.cn

## In Brief

COLD1 regulates G-protein signaling to confer chilling tolerance in rice, and a SNP in *COLD1* underlies the adaptation to cold environment in *japonica* rice.

## Highlights

- QTL *COLD1* regulates G-protein signaling to confer chilling tolerance in rice
- SNP2 in *COLD1*<sup>jap/ind</sup> enhances its ability to activate G-protein  $\alpha$  GTPase
- COLD1 interacts with G protein to activate the Ca<sup>2+</sup> channel for temperature sensing
- The SNP2-containing allele is selected during *japonica* rice domestication



# COLD1 Confers Chilling Tolerance in Rice

Yun Ma,<sup>1,7</sup> Xiaoyan Dai,<sup>1,7</sup> Yunyuan Xu,<sup>1,7</sup> Wei Luo,<sup>1,7</sup> Xiaoming Zheng,<sup>2</sup> Dali Zeng,<sup>3</sup> Yajun Pan,<sup>4</sup> Xiaoli Lin,<sup>1</sup> Huanhuan Liu,<sup>1</sup> Dajian Zhang,<sup>1</sup> Jun Xiao,<sup>1</sup> Xiaoyu Guo,<sup>1</sup> Shujuan Xu,<sup>1</sup> Yuda Niu,<sup>1</sup> Jingbo Jin,<sup>1</sup> Hui Zhang,<sup>1</sup> Xun Xu,<sup>5</sup> Legong Li,<sup>4</sup> Wen Wang,<sup>5</sup> Qian Qian,<sup>3</sup> Song Ge,<sup>2</sup> and Kang Chong<sup>1,6,\*</sup>

<sup>1</sup>Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

<sup>2</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

<sup>3</sup>State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China

<sup>4</sup>College of Life Science, Capital Normal University, Beijing 100048, China

<sup>5</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China

<sup>6</sup>National Center for Plant Gene Research, Beijing 100093, China

<sup>7</sup>Co-first author

\*Correspondence: [chongk@ibcas.ac.cn](mailto:chongk@ibcas.ac.cn)

<http://dx.doi.org/10.1016/j.cell.2015.01.046>

## SUMMARY

Rice is sensitive to cold and can be grown only in certain climate zones. Human selection of *japonica* rice has extended its growth zone to regions with lower temperature, while the molecular basis of this adaptation remains unknown. Here, we identify the quantitative trait locus *COLD1* that confers chilling tolerance in *japonica* rice. Overexpression of *COLD1<sup>jap</sup>* significantly enhances chilling tolerance, whereas rice lines with deficiency or downregulation of *COLD1<sup>jap</sup>* are sensitive to cold. *COLD1* encodes a regulator of G-protein signaling that localizes on plasma membrane and endoplasmic reticulum (ER). It interacts with the G-protein  $\alpha$  subunit to activate the  $\text{Ca}^{2+}$  channel for sensing low temperature and to accelerate G-protein GTPase activity. We further identify that a SNP in *COLD1*, SNP2, originated from Chinese *Oryza rufipogon*, is responsible for the ability of *COLD<sup>jap/ind</sup>* to confer chilling tolerance, supporting the importance of *COLD1* in plant adaptation.

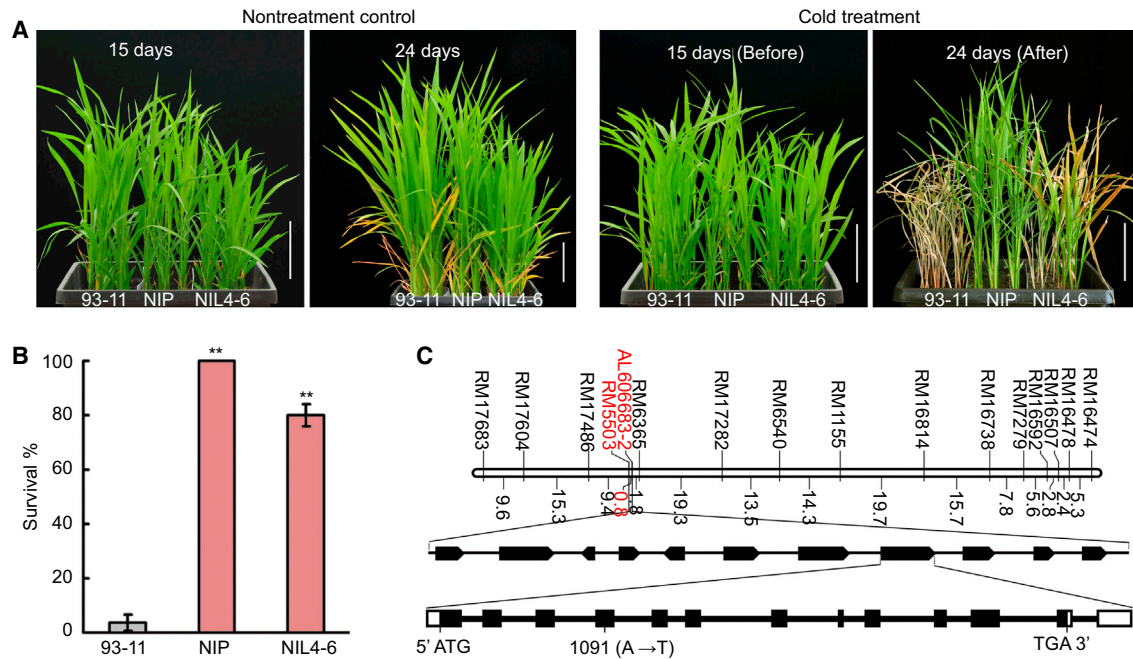
## INTRODUCTION

Rice, which is both a model plant and one that feeds more than half of the world's population (Sasaki and Burr, 2000), evolved in tropical and subtropical areas and is sensitive to chilling stress (Kovach et al., 2007; Saito et al., 2001; Sang and Ge, 2007). Extreme temperature thus represents a key factor limiting global rice plant distribution. Super hybrid rice cultivars produce high yields in tropical or subtropical climates but are frequently harmed by chilling. Therefore, molecular genetic tools have been urgently sought to improve rice chilling tolerance in order to maintain rice production in current regions and expand it into northern areas with lower yearly temperatures.

Asian cultivated rice (*Oryza sativa*) was domesticated from its wild relatives *Oryza nivara* and *O. rufipogon*. It consists of two

major subspecies, *indica* (*O. sativa* ssp. *indica*) and *japonica* (*O. sativa* ssp. *japonica*) (Kovach et al., 2007; Sang and Ge, 2007). Typical *japonica* cultivars, called *temperate japonica*, are grown in regions with lower yearly temperatures and generally exhibit stronger chilling tolerance than do *indica* cultivars. By contrast, some *japonica* cultivars that moved southwest to southeast Asia became tropical ecotypes, referred to as *javanica* or *tropical japonica*. Divergence between *indica* and *japonica* was driven by divergent natural selection imposed by contrasting environmental temperatures (Kovach et al., 2007; Sang and Ge, 2007). During human selection, cultivated rice has undergone significant changes in agricultural traits, such as grain yield, as well as environmental tolerance (Huang et al., 2012; Xu et al., 2012). Several developmental trait-related genes, such as *SH4* and *PROG1*, with signatures of domestication in cultivated rice have been identified using genetic mapping for quantitative trait loci (QTLs) and genome-wide association studies (GWAS) (Huang et al., 2012; Xu et al., 2012). The QTLs responsible for chilling tolerance in rice were mapped, revealing that the corresponding genes affect either seed germination or male sterility (Saito et al., 2001, 2010; Fujino et al., 2008; Koseki et al., 2010), but less is known about the molecular basis of the divergence between the two subspecies in terms of adaptation to the environment and geographical distribution.

Plant cellular adaptations to temperature differences are dependent on specific molecular cellular pathways including  $\text{Ca}^{2+}$ -mediated signal transduction. Cyclic nucleotide-gated channels (CNGCs) are nonspecific cation channels; in *Arabidopsis*, CNGCs form a family with 20 members and contribute to  $\text{Ca}^{2+}$  fluxes in various stress responses (Finka et al., 2012; Steinhilber and Kudla, 2013; Swarbreck et al., 2013). In mammals,  $\text{Ca}^{2+}$  channels interact with heterotrimeric guanine nucleotide-binding protein (G protein) complexes to function in stress responses (Wang and Chong, 2010). The transition of the mammalian G-protein  $\alpha$  subunit between an activated state and an inactivated is regulated by G-protein-coupled receptors (GPCRs), which mediate exchange (GDP release and GTP binding), and by regulator of G-protein signaling (RGS), which promotes GTP hydrolysis. Unlike animal G proteins, plant



**Figure 1. Map-Based Cloning of *COLD1***

(A) Phenotypic response to chilling in 93-11, Nipponbare (NIP), and the homozygote NIL4-6. Scale bars, 5 cm.

(B) The survival rate of 93-11, NIL4-6, and NIP after chilling treatment (96 hr). Values are expressed as mean  $\pm$  SD,  $n = 3$ ,  $**p < 0.01$ . See also Figure S1.

(C) The *COLD1* gene was mapped to the interval between the molecular markers AL606683-2 and RM5503 in chromosome 4. The gene was further delimited to a 77.33-kb genomic region on a BAC. Black arrows represent predicted genes. Black rectangles represent exons of *COLD1*. See also Table S1 and Figure S1.

heterotrimeric G proteins are self-activating and do not utilize GPCRs in converting to the GTP-binding state (Urano et al., 2013). Instead, the RGS with activity of GTPase-accelerating protein (GAP) activity for GTP hydrolysis is more important for G-protein signaling in plant cells. In response to mild heating shock,  $\text{Ca}^{2+}$ -permeable channels mediate signals that lead to an influx of  $\text{Ca}^{2+}$  into plant cells (Saidi et al., 2009).  $\text{Ca}^{2+}$  signaling in plant cells also occurs during cold shock (Knight et al., 1996), although less is known about how the cold shock is linked to  $\text{Ca}^{2+}$  signaling. Overall, it is well established that  $\text{Ca}^{2+}$  signaling pathways and the resultant changes in gene transcription are involved in responses to altered temperature in plant cells (Dai et al., 2007; Lee et al., 2009; Ma et al., 2009). However, it is unknown how the signaling pathway in response to cold stimulation evolved during the divergence between rice subspecies *indica* and *japonica*.

Here, we provide evidence that a QTL gene, *CHILLING-TOLERANCE DIVERGENCE 1* (*COLD1*), is associated with divergence in chilling tolerance of rice cultivars. We further demonstrate that a single-nucleotide mutation at *COLD1* confers adaptation of *japonica* rice to chilling and originated from the Chinese wild populations of *O. rufipogon*. *COLD1* localized at the plasma membrane, and endoplasmic reticulum (ER) is involved in sensing cold to trigger  $\text{Ca}^{2+}$  signaling for chilling tolerance. These findings reveal the importance of *COLD1* in plant adaptation and its great potential for rice molecular breeding.

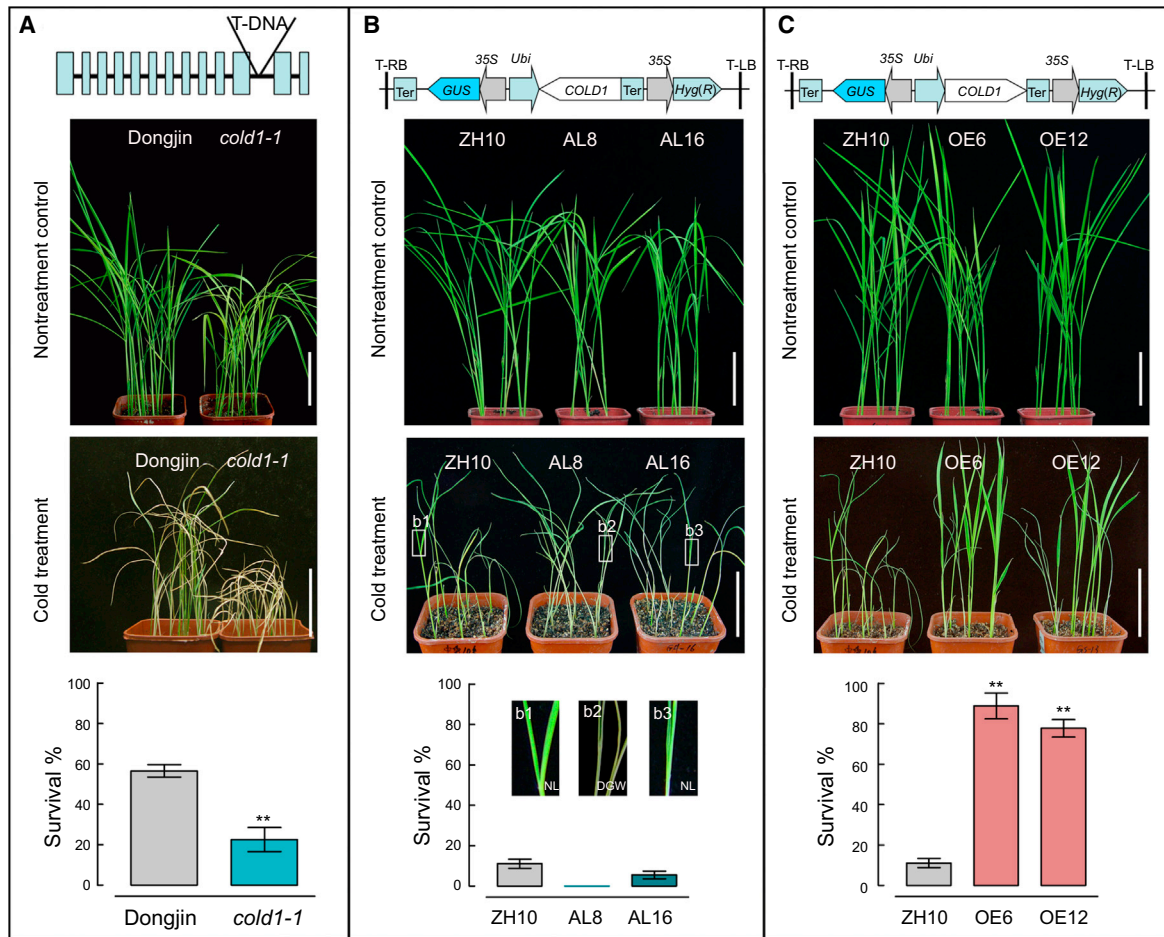
## RESULTS

### *COLD1* Confers Chilling Tolerance in Rice

Chilling tolerance of rice cultivars is regulated by QTLs derived from the subspecies *japonica* (Saito et al., 2001). To identify the genes involved in the increased chilling tolerance found in cultivars from growth regions with low yearly temperatures, we carried out a QTL analysis for *chilling-tolerance divergence* (*COLD*) in recombinant inbred lines (RILs) generated from a cross between chilling-tolerant Nipponbare (*japonica*) and chilling-sensitive 93-11 (*indica*) cultivars, testing for chilling sensitivity using the cold treatment ( $4^{\circ}\text{C}$ ) (Figure 1A). Using 151 RILs, we detected five QTLs, on chromosomes 1, 2, 4, 6, and 8 (Table S1). One of them, *COLD1*, was defined between markers RM6365 and RM5503 on the long arm of chromosome 4 (Figure 1C; Table S1). This locus explained 7.23% of the variance in chilling tolerance and shared the same locus with the QTL *Ctb2* despite slight differences in the crossed populations (Saito et al., 2001). The *COLD1* locus displayed much lower interaction with other QTLs for chilling tolerance ( $p = 0.0363$ ,  $0.0242$ ) than did the other loci, such as *COLD4* ( $p = 0.0002$ ) and *COLD5* ( $p = 0.0006$ ) (Table S1).

To evaluate whether the Nipponbare (NIP) locus, *COLD1*<sup>NIP</sup>, contributes to chilling tolerance, we generated three near-isogenic lines (NILs) containing the *COLD1*<sup>NIP</sup> locus in the 93-11 genetic background, which is one of the parental lines of the Chinese super hybrid rice. The homozygous *COLD1*<sup>NIP/NIP</sup> lines NIL4-1 and NIL4-6 showed remarkably higher tolerance





**Figure 2. *COLD1* Is Essential for Chilling Tolerance**

(A) The *cold1-1* mutant showed chilling sensitivity. The survival rate was determined after treatment at 4°C for 96 hr and subsequent recovery at 30°C for 7 days. (B) The antisense transgenic rice lines (AL8 and AL16) showed chilling sensitivity. The survival rate was determined after treatment at 2°C–3°C for 96 hr and subsequent recovery at 30°C for 4 days. Panes are enlargements of plants showing live seedlings with new leaves (NL) and dead seedlings with dry green and white leaves (DGW).

(C) The overexpression transgenic lines (OE6 and OE12) showed chilling tolerance. The survival rate was determined after treatment at 2°C–3°C for 96 hr and subsequent recovery at 30°C for 4 days. The upper diagrams represent the T-DNA insertion or the transgenes used to generate the lines.

35S, CaV 35S promoter; Ubi, maize ubiquitin promoter; T-RB, T-DNA, right border; T-LB, T-DNA, left border; GUS,  $\beta$ -glucuronidase; Hyg (R), Hygromycin B resistance, Ter, terminator. Values are means  $\pm$  SD,  $n = 3$ . Scale bars, 5 cm. \*\* $p < 0.01$ . See also Figures S2 and S3.

to chilling compared to 93-11 (Figures 1B and S1). A dominance assay on the heterozygote *COLD1*<sup>NIP/93-11</sup> NIL2-5 showed that its chilling tolerance was similar to that of NIL4-1 and NIL4-6 (Figure S1). To fine-map *COLD1*, we analyzed 8,368 F<sub>2</sub> plants generated from NIL2-5 and narrowed the candidate region to 77.33 kb between AL606683-2 and RM5503. This region contains 11 predicted genes or open reading frames (Figure 1C; Table S1). Genomic DNA sequence comparisons between the candidate regions of the parents NIP and 93-11 showed that one single-nucleotide mutation at 15<sup>th</sup> nucleotide in the fourth exon of *COLD1* (A in NIP was changed into T in 93-11) (LOC\_Os04 g51180, MSU Rice Genome Annotation (Osa1) release 7. <http://rice.plantbiology.msu.edu>) caused a change in an encoded amino acid (Lys in NIP was changed into Met in 93-11) (Figure 1).

To determine whether the *COLD1* gene underlies the QTL, we constructed *COLD1*<sup>ja</sup>-overexpression (OE) and antisense (AL) transgenic rice lines in *japonica* cultivar Zhonghua 10 (ZH10) (Figures 2 and S2), and examined their chilling tolerance. In addition, we analyzed the *cold1-1* mutant, which has a T-DNA insertion in the 11<sup>th</sup> intron of *COLD1*, +3,707 bp downstream from the ATG in the *japonica* rice Dongjin (DJ) background, and which lacks the full-length transcript (Figure S2). Seedlings were exposed to chilling temperature (4°C) and subsequently returned to 30°C. Rice plants with chilling tolerance were defined as those that could re-differentiate new leaves or continue growing leaves when returned normal conditions after treatment with chilling stress. Clear phenotypic differences in the survival rate (percentage alive seedlings of the total tested plants) were observed among these lines (Figures 2 and S2). Seedlings of the *cold1-1*

mutant, as well as of the antisense lines (AL5, AL6, AL8, and AL16) were chilling sensitive compared to the wild-type (WT). By contrast, *COLD1*<sup>jap</sup>-overexpression lines, such as OE6, OE12, OE1, and OE2, showed higher chilling tolerance than WT. The findings suggest that *COLD1* modulates chilling tolerance in rice.

### SNP2 Is Associated with Chilling Tolerance

To test for association between *COLD1* alleles and chilling tolerance, we examined the chilling tolerance of 5 *indica* and 20 *japonica* cultivars, as well as 2 accessions of wild rice (Table S2). All *japonica* cultivars and 2 *O. rufipogon* accessions showed stronger chilling tolerance than did all *indica* cultivars (Figure 3A; Table S2). We then sequenced the full-length *COLD1* gene of 4.78 kb including the 5' and 3' untranslated regions in these samples and identified seven SNPs (Figure 3A), including a synonymous polymorphism in the first exon (SNP1), a nonsynonymous polymorphism only in the fourth exon (SNP2), and five substitutions in introns (SNP3, 4, 5, 6, and 7). We grouped the cultivars based on chilling sensitivity and examined whether chilling tolerance was associated with allelic differences (SNPs) in *COLD1*. Strikingly, all accessions with confirmed chilling tolerance, including 20 *japonica* cultivars and 2 *O. rufipogon* accessions, differed from the *indica* cultivars that lacked the chilling tolerance by the SNP in the fourth exon (SNP2). The nucleotide polymorphism of T/C versus A in the fourth exon resulted in Met<sup>187</sup>/Thr<sup>187</sup> in *indica* compared to Lys<sup>187</sup> in *japonica* cultivars. At the remaining SNP sites, polymorphic nucleotides were found in cultivars both with and without chilling tolerance (Figure 3A and Table S3).

To determine whether SNP2 led to alteration of chilling tolerance, we generated transgenic lines overexpressing the gene from *indica* plants (SNP2<sup>ind(T)</sup>) in the *japonica* ZH11 background (Figures 3 and S3). The *COLD1*<sup>ind</sup> transgenic lines were more sensitive to chilling compared to ZH11. In addition, the transgenic lines of *COLD1*<sup>ind</sup> (SNP2<sup>ind(T)</sup>) in the *cold1-1* mutant background showed a similar chilling tolerance as *cold1-1*, but significantly weaker tolerance than wild-type DJ. By contrast, the transgenic lines of *COLD1*<sup>jap</sup> (SNP2<sup>jap(A)</sup>) in the *cold1-1* background showed similar tolerance as wild-type after cold treatment (Figures 3B and S3). Together with the enhanced chilling tolerance observed in the *COLD1*<sup>jap</sup> (SNP2<sup>jap(A)</sup>) transgenic lines in wild-type background and that in the *cold1-1* background for the genetic complementation (Figure 2C), this suggests that SNP2, resulting in a change of encoded amino acid, is responsible for chilling tolerance in *japonica* rice.

### SNP2 Arose during *japonica* Domestication

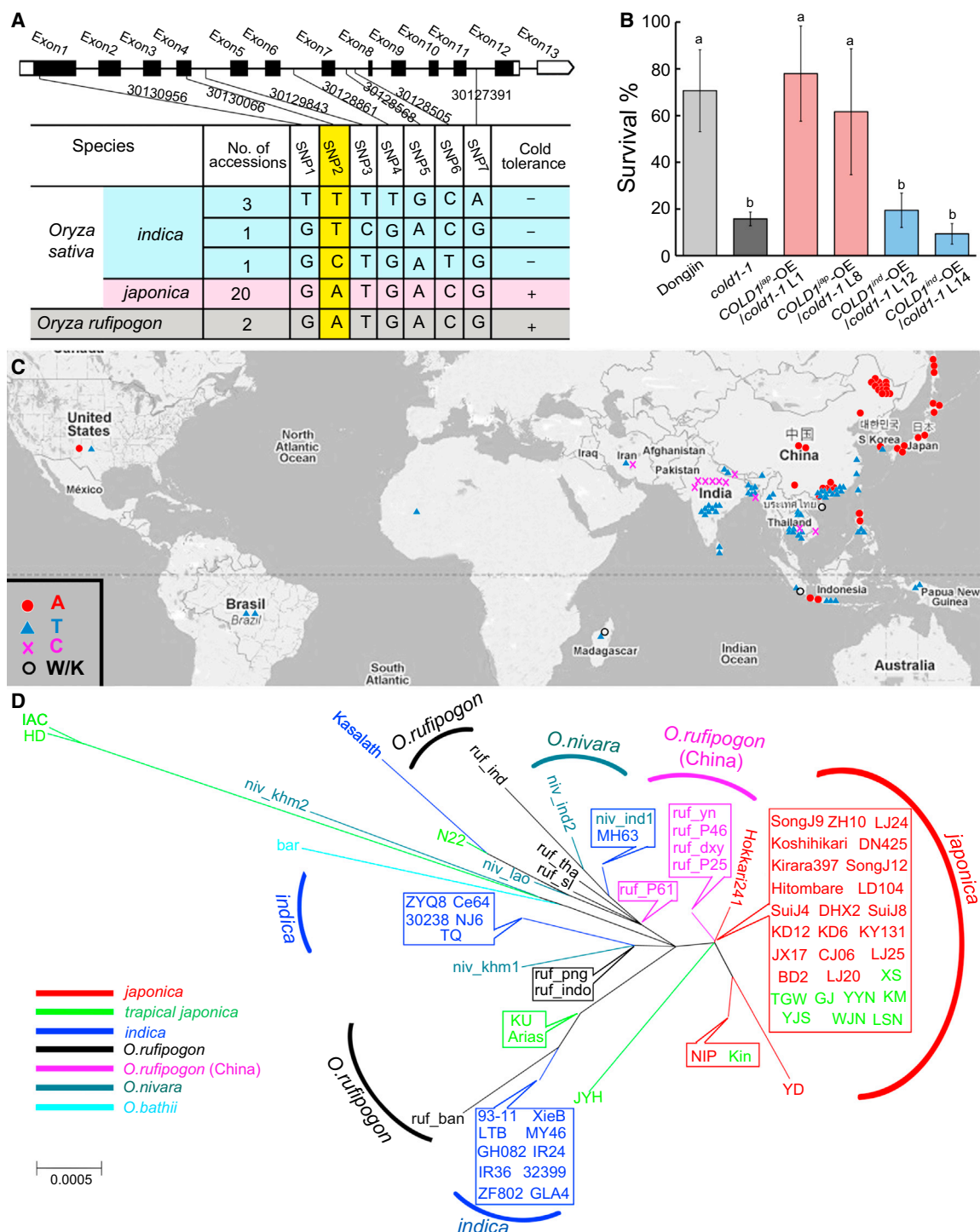
To examine the evolutionary origin of the alleles, we sequenced the full-length *COLD1* gene in an additional 100 accessions of cultivated and wild rice, including 36 *indica*, 15 *japonica*, and 15 *javanica* accessions, and 14 *O. nivara* and 19 *O. rufipogon* individuals as well as one *O. barthii* individual (Table S2). All *japonica* accessions, except for two samples displaying heterozygosity, had nucleotide A at the SNP2 site, whereas the *indica* accessions had either T or C, and *javanica* had A or T or C at this site. The five *O. rufipogon* samples originated from China

had A at this site, and one *O. rufipogon* sample from Hainan province in China had W, whereas the remaining wild rice samples including 15 *O. rufipogon* samples from outside of China, 14 *O. nivara* samples and one *O. barthii* sample had either T or C (Table S2).

Geographically, 33 *japonica* cultivars, one *javanica* and the Chinese *O. rufipogon* samples with A at SNP2 were distributed in the northern area of China, Japan, Korea, and the United States, or at higher elevations of the southeast zone of Asia (Figure 3C). By contrast, all samples without A at SNP2, including 41 *indica* and 15 *O. rufipogon* samples from outside of China, were distributed in southern and southeastern Asia, regions with higher yearly temperatures. For *javanica*, 14 samples with nucleotide diversities at the site were distributed in regions of higher yearly temperature, such as southern area of China and the Philippines. Phylogenetic analysis of the *COLD1* sequences of the 72 accessions sampled (Table S2) indicated that all *japonica* accessions and the Chinese *O. rufipogon* samples carrying the chilling-tolerance SNP2<sup>A</sup> were grouped together with 60% bootstrap support (Figure 3D). These observations indicate that the *COLD1* allele with the mutation at SNP2<sup>A</sup> is likely to have originated from Chinese *O. rufipogon* during *japonica* rice domestication.

To examine whether selection has acted on *COLD1*, we analyzed nucleotide diversity across the sequenced region in 72 accessions (Table S2), including the original 27 accessions tested for chilling tolerance. A comparison of the nucleotide diversity among *indica*, *japonica*, *javanica*, *O. nivara*, and *O. rufipogon* indicated that on average, *japonica* exhibited much lower diversity ( $\theta = 0.0004$ ;  $\pi = 0.0002$ ) than *indica* ( $\theta = 0.0014$ ;  $\pi = 0.0013$ ), *javanica* ( $\theta = 0.0025$ ;  $\pi = 0.0017$ ), and the two wild rice species ( $\theta = 0.0014$ – $0.0022$ ;  $\pi = 0.0010$ – $0.0020$ ). Significantly negative Tajima's *D* values were observed only for *japonica* cultivars (Table S3), consistent with selection at the *COLD1* locus.

To determine further whether the reduction in nucleotide diversity in *japonica* rice could be caused by artificial selection, we conducted MLHKA tests on *COLD1* sequences for all six taxa (Table S3) in reference to seven neutral genes (Zhu et al., 2007). We found a significant value for *japonica* rice ( $p = 0.001$ ), indicative of strong artificial selection on the *COLD1* locus during *japonica* domestication. To exclude the potential impact of demography on diversity reduction at *COLD1*, we further examined the nucleotide diversity for the ten genes within 400-kb region surrounding the *COLD1* locus in 43 accessions (Tables S2 and S3) because selection might lead to a selective sweep in the flanking region of the selected genes (Asano et al., 2011). As expected, we found that the average nucleotide diversity of the ten genes in *japonica* ( $\pi = 0.0003$ ) was much lower than those of all other rice groups ( $\pi = 0.0027$  for *indica*;  $\pi = 0.0020$  for *javanica*;  $\pi = 0.0057$  for wild rice) (Table S3), consistent with the selective sweep argument. A coalescent simulation using the ten surrounding genes revealed a significant lower *K* value (the severity of the bottleneck) in *japonica* ( $K = 0.06$ ) than that of neutral genes ( $K = 0.2$ ) ( $p = 0.0097$ ) (Table S3) (Zhu et al., 2007), indicating that the reduced diversity at the genes surrounding *COLD1* in *japonica* cannot be explained by a domestication bottleneck alone. Taken together, our data



**Figure 3. Association of SNPs in *COLD1* with Chilling Tolerance and Their Geographic and Phylogenetic Origins**

(A) SNPs and chilling tolerance in 27 accessions.

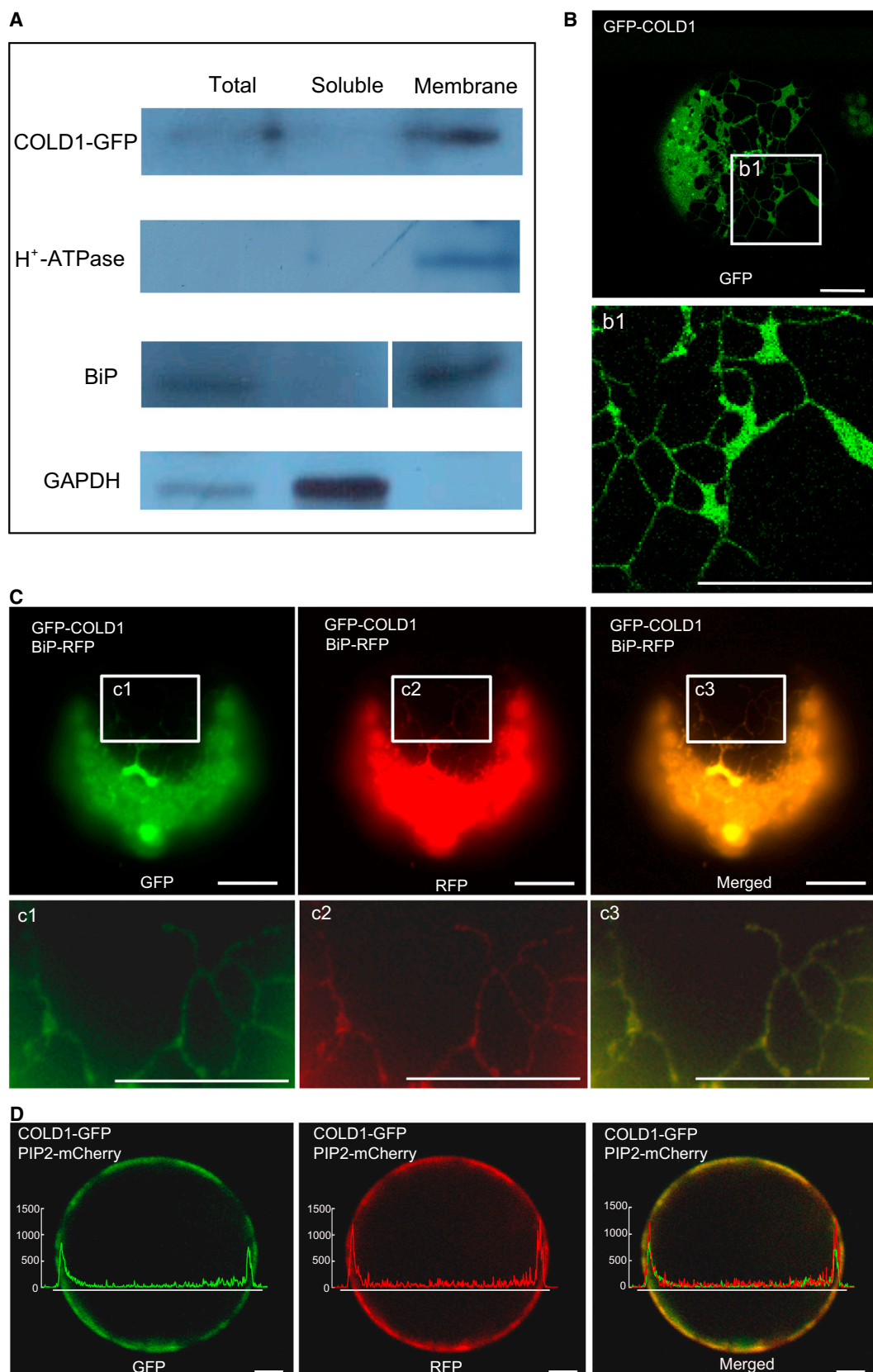
(B) Chilling tolerance response of *COLD1* complementation lines in the *cold1-1* genetic background. Values are expressed as means  $\pm$  SD,  $n = 3$ . Statistically different values ( $p < 0.05$ ) are indicated by different letters.

(C) Geographic distribution of 127 accessions tested (Table S2). The *japonica* and *O. rufipogon* samples carrying A at the SNP2 site are represented by red circles. The *indica* cultivars with T/C are denoted by blue triangles/purple crosses, respectively. The heterozygous cultivars [W (A or T)/K (G or T)] are represented by black rings.

(D) Neighbor-joining tree. Bootstrap values over 60% are given on the branches.

See also Tables S2 and S3.





(legend on next page)



show that the A at the functional SNP2 of *COLD1* is associated with the development of chilling tolerance in cultivated rice and might represent an ancient allele preserved in the Chinese populations of *O. rufipogon* and selected during domestication of *japonica* rice.

### COLD1 Localizes to the ER and Plasma Membrane

*COLD1* was predicted to encode a 53-kDa protein with nine transmembrane domains. As expected, it was grouped with its orthologs from the monocotyledons in a phylogenetic tree (Figure S4). Immunoblotting assays on tissues expressing a *COLD1-GFP* fusion transgene showed signal from an anti-GFP antibody only in the membrane protein fraction, similar to the control membrane proteins H<sup>+</sup>-ATPase and BiP, a marker of the endoplasmic reticulum (ER). No signal for *COLD1-GFP* was found in the soluble fractions, although the soluble control of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) protein did show a signal (Figure 4A). Under microscopy, fluorescence of GFP-*COLD1* overlapped with that of BiP-RFP at the ER (Figures 4B, 4C, and S4D) and with that of PIP2-mCherry, a marker for the plasma membrane and ER (Lee et al., 2009), at the plasma membrane (Figure S4E). Similarly, the signal of *COLD1-GFP* co-localized with that of PIP2-mCherry at ER with a reticular pattern and at the plasma membrane (Figures S4F, S4G, and 4D). The plasma membrane localization was independent on the myristoylation of G2 in the N-terminal motif M1-G2-W3 of *COLD1* (Figures S4H and S4I) (Batistic et al., 2008; Yamauchi et al., 2010). These results suggest that *COLD1* is mainly localized to the ER and plasma membrane.

### COLD1 Interacts with G-Protein $\alpha$ Subunit

Based on hidden Markov model (Krogh et al., 2001) predictions, *COLD1* contains nine transmembrane domains with a preferred orientation of an extracellular N terminus and an intracellular terminus, similar to the pattern of its *Arabidopsis* orthologs (Figures S4 and S5), GTG1/2, which interact with G-protein  $\alpha$  subunit. We confirmed the interaction between *COLD1* and the rice G-protein  $\alpha$  subunit 1 (RGA1) (Ludewig et al., 2003; Stagljar et al., 1998) in vitro and in vivo. Yeast cells co-transformed either with *COLD1<sup>jap</sup>* or *COLD1<sup>ind</sup>* or *COLD1<sup>Δjap</sup>* and RGA1 grew well on medium lacking His and Ade and showed X-gal staining, in contrast to the negative controls (Figure S5). In co-immunoprecipitation (Co-IP) assays, GFP-*COLD1* was detected in complexes immunoprecipitated with the anti-FLAG antibody from leaves of transgenic plants expressing GFP-*COLD1* and FLAG-RGA1 (Figure 5A). Bimolecular fluorescence complementation (BiFC) assays revealed reconstituted YFP fluorescence in the

plasma membrane of transgenic lines harboring *COLD1-YFP<sup>C</sup>* and RGA1-YFP<sup>N</sup> (Figure 5B). By contrast, no fluorescence was detected in the negative controls OsBAK1-YFP<sup>C</sup> and RGA1-YFP<sup>N</sup>. These data demonstrate that *COLD1* can physically interact with RGA1 in plant cells.

### COLD1 Functions as a GTPase-Accelerating Factor on RGA1

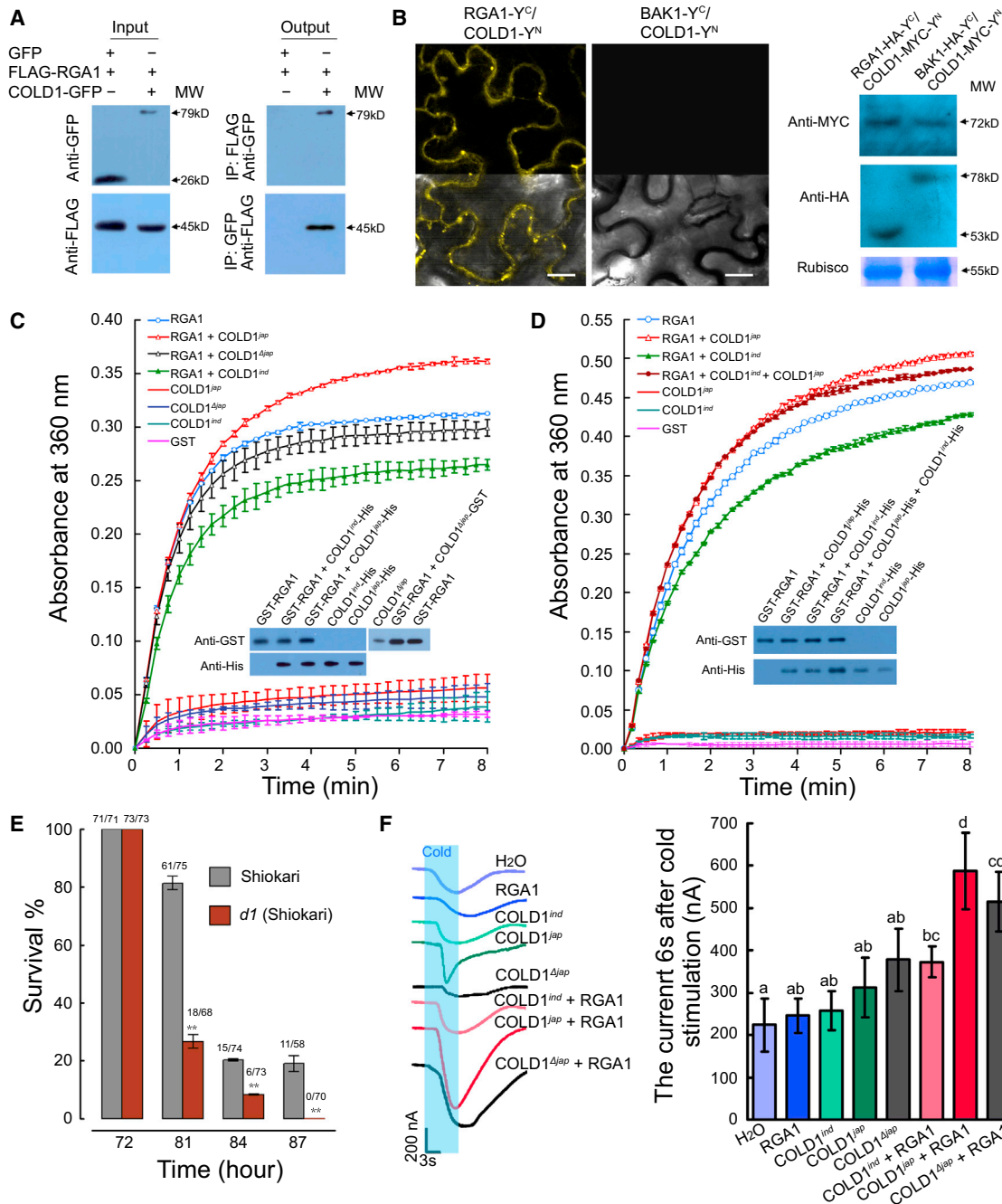
Biochemical activity assays confirmed that RGA1 instead of *COLD1* alone had GTPase activity, dependent on Mg<sup>2+</sup> concentration in the reaction (Figures 5C, 5D, and S5D). RGA1 GTPase activity was accelerated in the presence of *COLD1<sup>jap</sup>* (SNP<sup>jap(A)</sup>). By contrast, *COLD1<sup>ind</sup>* (SNP<sup>ind(A)</sup>), as well as the truncated protein *COLD1<sup>Δjap</sup>* from *cold1-1*, suppressed RGA1 GTPase activity over the course of the assay (Figure 5C). The *COLD1<sup>jap</sup>*-induced acceleration of RGA1 GTPase activity was impaired by inclusion of *COLD1<sup>ind</sup>* in the reaction (Figure 5D), which may explain the tolerance differences between *COLD1<sup>ind</sup>* and *COLD1<sup>jap</sup>* transgenic lines on the *japonica* background, as well decreased tolerance of *cold1-1* (Figure S3). A time-course assay for the tolerance showed that the RGA1 mutant *d1* was significantly more sensitive to chilling for survival compared with wild-type Shiokari (Figure 5E). This is consistent with that the *COLD1* and RGA1 complex is required for the tolerance.

We used an electrode voltage clamp approach to record the currents of oocytes co-expressing *COLD1* and RGA1 (Figure 5F). Upon cold treatment, an inward current was significantly activated in the cells co-expressing *COLD1<sup>jap</sup>* and RGA1 compared with expression alone, which was in contrast to their patterns showing no response to heat stimulation (40°C) (Figure S5) (Finka et al., 2012). The cold-activated response lagged by several seconds and returned rapidly to baseline levels after removal of cold stimulation. The cold-stimulated inward current was 588 ± 90 nA. By contrast, control cells and oocytes co-expressing *COLD1<sup>ind</sup>* and RGA1 generated background currents of 373 ± 36 and 246 ± 41 nA, respectively. Co-expression of the truncated gene *COLD1<sup>Δjap</sup>* and RGA1 led to a weaker inward current in response to cold stimulation than that of *COLD1<sup>jap</sup>*. This suggests that the cold-stimulated inward current signal is dependent on interaction between *COLD1* and RGA1 in the presence of Ca<sup>2+</sup>. Probably, a complex of *COLD1* that has a GTPase-accelerating on RGA1 may affect influx of cations (such as Ca<sup>2+</sup>) to cause changes of the membrane currents in oocyte cells. The *japonica* allele *COLD1<sup>jap</sup>* showed a stronger response with RGA1 on the cold-stimulated inward current signal than did the *indica* allele *COLD1<sup>ind</sup>*.

### Figure 4. COLD1 Localization

(A) Immunoblotting assay showing GFP antibody recognized GFP-tagged *COLD1* in the membrane protein fraction from transgenic tobacco. H<sup>+</sup>-ATPase, membrane protein control; BiP, ER marker control; GAPDH, glyceraldehydes-3-phosphate hydrogenase soluble protein control.  
(B) ER localization of *COLD1* in *Arabidopsis* protoplast cells. The b1 images (lower) show enlargements of the regions framed in white (upper).  
(C) Co-localization of *COLD1* with ER marker. GFP-*COLD1* signal was merged with that of the RFP-tagged BiP marker in *Arabidopsis* mesophyll protoplasts. The images with labels c1, c2, and c3 (lower) are enlargements of the regions framed in white (upper). Scale bars, 10 μm.  
(D) Plasma membrane localization of *COLD1* in cells. *COLD1-GFP* signal was merged with that of the PIP2-mCherry (an intrinsic plasma membrane protein) marker in Tobacco mesophyll protoplasts. The fluorescence intensity was scanned with the ImageJ plot profile tool (ImageJ v.1.47; <http://rsbweb.nih.gov/ij/download.html>). y axes are relative pixel intensity. Scale bar, 10 μm.

All experiments were performed with at least three biological replicates. See also Figure S4.



**Figure 5. COLD1 Interacts with RGA1**

(A) Co-immunoprecipitation assays confirming the interaction between COLD1 and RGA1. Co-expressed FLAG-RGA1 and COLD1-GFP in tobacco leaves were immunoprecipitated by anti-FLAG or -GFP. Blots were probed with anti-GFP or -FLAG.

(B) BiFC assays showing that the proteins interact in vivo. The bottom ones are the merged images. Immunoblots (right) confirmed the expression of the interaction proteins in the transgenic leaf tissues used in the BiFC assay. Y<sup>N</sup>, YN173; Y<sup>C</sup>, YCM. Scale bars, 20  $\mu$ m.

(C) Intrinsic GTPase activity of RGA1 was accelerated by COLD1<sup>Δiap</sup> but impaired by COLD1<sup>ind</sup> or COLD1<sup>Δiap</sup>. The molar ratio of RGA1/COLD1 was 4.8. Values are expressed as mean  $\pm$  SD, n = 3. The immunoblots show amount of proteins in the reaction.

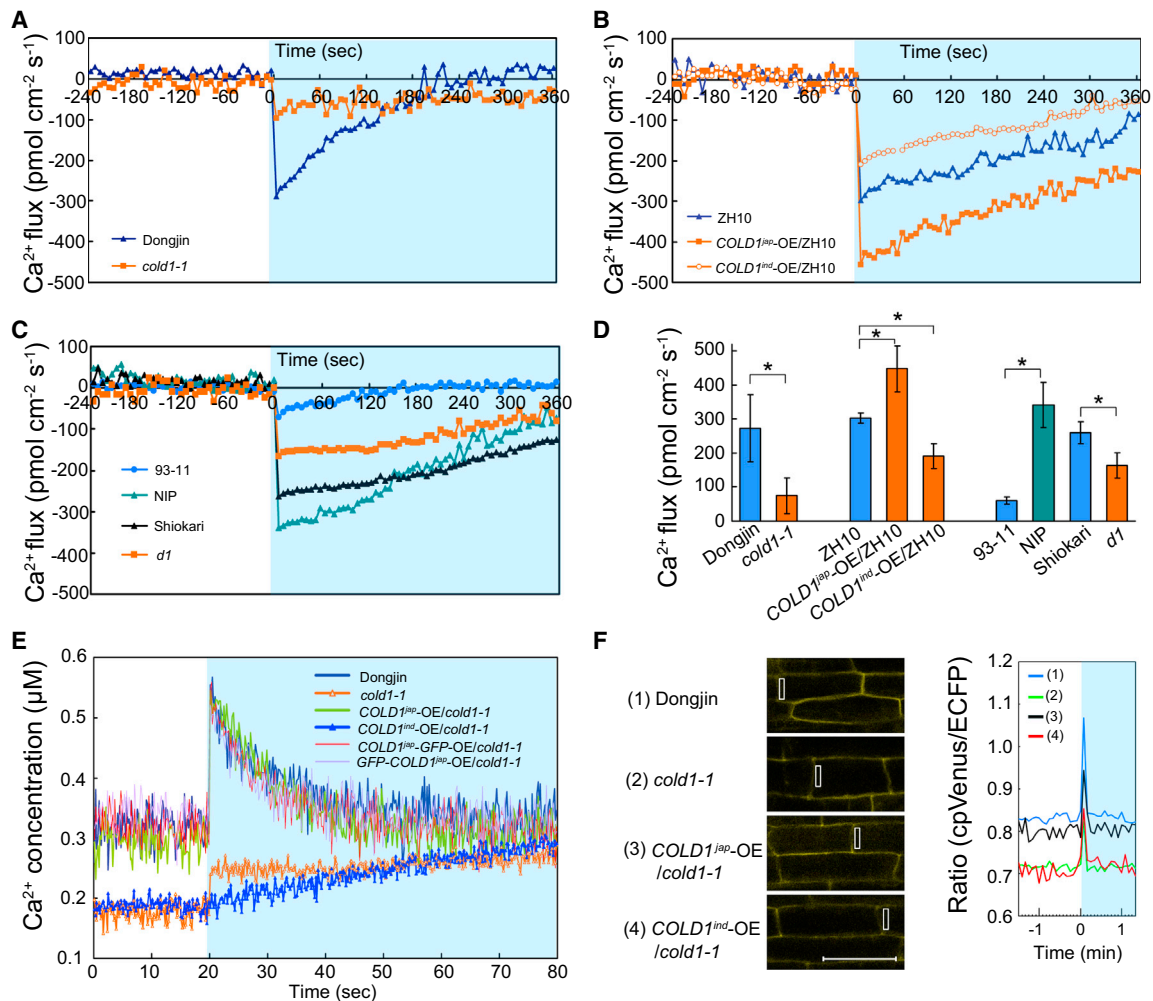
(D) Acceleration of RGA1 GTPase activity by COLD1<sup>Δiap</sup> was inhibited by addition of COLD1<sup>ind</sup> in vitro. The molar ratio of RGA1/COLD1 was 4.8. Values are expressed as mean  $\pm$  SD, n = 3. The immunoblots show amount of proteins in the reaction.

(E) Time course of chilling tolerance showing that the d1 mutant is sensitive to cold treatment. The numbers above the bars are alive and total plants. Values are expressed as mean  $\pm$  SD, n = 3; \*\*p < 0.01.

(F) Electrophysiological characterization of *Xenopus* oocytes co-expressing COLD1 and RGA1, as well as the control RGA1 only. The blue background represents a duration for cold treatment in solution. The holding potential was -110 mV.

Values are expressed as means  $\pm$  SD, n = 7. Statistically different values (p < 0.05) are indicated by different letters. See also Figure S5.

(legend continued on next page)



**Figure 6.  $\text{Ca}^{2+}$  Signaling upon Cold Shock in Rice Plants**

(A–C) SIET measurements show extracellular  $\text{Ca}^{2+}$  influx upon cold shock in live roots of various genetic backgrounds ( $n > 6$ ).

(D) Significance testing of the mean maximal  $\text{Ca}^{2+}$  influxes. Values are expressed as mean  $\pm$  SD,  $n > 6$ , Student's *t* test, \* $p < 0.05$ .

(E)  $[\text{Ca}^{2+}]_{\text{cyt}}$  monitored with aequorin in response to cold shock in wild-type Dongjin and the *cold1-1* mutant ( $n > 6$ ).

(F) Cold response of  $[\text{Ca}^{2+}]_{\text{cyt}}$  in live root cells using Yellow Cameleon (NES-YC3.6). Scale bars, 50  $\mu\text{m}$ . The rectangles represent regions of interest (ROIs) considered for ratiometric measurements. The numbers used for ratiometric measurements are indicated in the boxes. The experiments were replicated at least three times. The blue background represents a duration for cold treatment.

See also Figure S6.

### **COLD1 Is Essential for Changes in $\text{Ca}^{2+}$ Influx upon Cold Treatment**

To examine  $\text{Ca}^{2+}$  flux in response to cold shock, we used the scanning ion-selective electrode technique (SIET) on rice roots (Ludewig et al., 2003). Upon cold stimulation, there was a significant influx of extracellular  $\text{Ca}^{2+}$  with a minus peak in wild-type Dongjin roots (Figures 6A and S6). By contrast, *cold1-1* showed no remarkable changes in SIET signals under the same conditions. Compared with wild-type ZH10, the *COLD1<sup>ind</sup>-OE* transgenic line exhibited more  $\text{Ca}^{2+}$  influx in response to cold treatment, but the *COLD1<sup>ind</sup>-OE* transgenic line displayed less (Figure 6B). Nipponbare, *japonica* rice, showed a stronger response than did *indica* 93-11 (Figure 6C). In addition, the *d1* mutant of *RGA1* showed less  $\text{Ca}^{2+}$  influx than did wild-type Shikari. The mean

maximal influxes of cold shock between *cold1-1* or transgenic lines and wild-type were significantly different (Figure 6D). In response to salt stress, by contrast, the overlapped SIET patterns between *cold1-1* and DJ indicated that salt stimulation signaling may be independent to *COLD1* (Figure S6). The extracellular  $\text{Ca}^{2+}$  influx peaks in response to cold shock hint that the net cytoplasm  $[\text{Ca}^{2+}]_{\text{cyt}}$  derived from bulk extracellular  $\text{Ca}^{2+}$  might be substantially increased.

We also monitored  $\text{Ca}^{2+}$  concentration in the cytoplasm ( $[\text{Ca}^{2+}]_{\text{cyt}}$ ) using cytosolic aequorin. Immediately upon the onset of cold treatment, Dongjin showed a significant  $[\text{Ca}^{2+}]_{\text{cyt}}$  peak up to  $0.554 \pm 0.013 \mu\text{M}$  from  $0.319 \pm 0.029 \mu\text{M}$  ( $n = 7$ ), which then decreased (Figure 6E). By contrast, *cold1-1* showed a much smaller increase in  $[\text{Ca}^{2+}]_{\text{cyt}}$  from  $0.177 \pm 0.014$  to

$0.240 \pm 0.040 \mu\text{M}$  ( $n = 9$ ) and subsequently maintained a nearly stable level under the same conditions (Figure 6E). With regard to calcium level, the cold shock pattern of  $[\text{Ca}^{2+}]_{\text{cyt}}$  in the *COLD1<sup>jap</sup>*-complemented lines (harboring either *COLD1<sup>jap</sup>*-GFP or GFP-*COLD1<sup>jap</sup>*) ( $0.545 \pm 0.042 \mu\text{M}$  [ $n = 6$ ]) nearly overlapped with that of wild-type, whereas the *COLD1<sup>ind</sup>* transgenic line on *cold1-1* ( $0.186 \pm 0.011 \mu\text{M}$  [ $n = 6$ ]) showed similar pattern as *cold1-1* (Figure 6E).

We used the Cameleon technique to further confirm the genetic complementation effect on  $\text{Ca}^{2+}$  elevation (Krebs et al., 2012). The root cells of DJ showed a remarkable cytoplasm  $\text{Ca}^{2+}$  peak after cold treatment, while *cold1-1* had a weaker peak, as well as a relatively low basal level (Figure 6F). The complemented lines of *COLD1<sup>jap</sup>* almost completely rescued the cold-stimulated  $\text{Ca}^{2+}$  elevation in the *cold1-1* background. It is also notable that the recovered  $\text{Ca}^{2+}$  patterns of *COLD1<sup>jap</sup>* included the basal elevation compared to *cold1-1*. By contrast, overexpression *COLD1<sup>ind</sup>* in *cold1-1* did not rescue  $\text{Ca}^{2+}$  response in either the peak or basal level. In addition, the trends on fluorescence dyeing data for  $[\text{Ca}^{2+}]_{\text{cyt}}$  cold responses were in accord with these results (Figure S6).

In addition, the genetic complementation lines of *COLD1<sup>jap</sup>* in *cold1-1* background showed more remarkable cold-induced expression patterns for the stress-specific downstream genes, such as *OsAP2*, *OsDREB1A*, *OsDREB1B*, and *OsDREB1C* than did the overexpression of *COLD1<sup>ind</sup>* line (Figure S6). Thus, the findings on both the extracellular  $\text{Ca}^{2+}$  influx and the net cytoplasm  $[\text{Ca}^{2+}]_{\text{cyt}}$  signaling are consistent with the idea that *COLD1* is essential for cold shock-dependent intracellular  $\text{Ca}^{2+}$  changes in rice.

## DISCUSSION

In this work, we identified the QTL *COLD1*, which is required for chilling tolerance in *japonica* rice during the seedling stage. The *COLD1* locus enhanced chilling tolerance in near-isogenic lines NIL4-1 and NIL4-6 from the background cultivar *indica* 93-11 (Figure S1). It is worth noting that mature rice plants of both NILs with chilling tolerance displayed increased seed number per panicle and maintained grain yield per plant compared with 93-11, which is one of the desirable parental lines of the Chinese super hybrid rice. Thus, these NILs could potentially be used as parents of super hybrid rice, conferring chilling tolerance without negative effects on grain yield. This finding, along with the enhanced tolerance of the *COLD1<sup>jap</sup>* overexpression lines, emphasizes the potential of either genetic or transgenic approaches to improve chilling tolerance for rice breeding.

Chilling tolerance, i.e., the capacity to reestablish differentiation and growth under normal conditions after cold exposure, is a complex trait in seedlings that is controlled by multiple QTLs. Most of the QTLs genetically interacted with each other, resulting in a higher genetic contribution to chilling tolerance in the population. For instance, the *COLD2* QTL interacted genetically either with *COLD4* or *COLD5* resulting in an overall contribution to chilling tolerance of more than 16.8% (Table S1). By contrast, *COLD1* did not genetically interact with other QTLs and already alone contributed 7.23% to overall chilling tolerance. Nucleotide diversity analysis suggested that there was

strong artificial selection on the *COLD1* locus during *japonica* domestication (Tables S2 and S3).

*COLD1*'s topology, localization and interaction with RGA1, as well as its regulatory effects on RGA1 GTPase activity, support the idea that *COLD1* is a RGS with GTPase-accelerating activity, similar to AtRGS1 (Chen et al., 2003; Johnston et al., 2007; Shabala and Newman., 2000; Stagljar et al., 1998; Urano et al., 2012). The subcellular localization pattern of *COLD1* on the ER and plasma membrane partially overlaps those of its *Arabidopsis* orthologs GTG1/2 (Johnston et al., 2007; Pandey et al., 2009), but *COLD1* is different from those GTG1/2 in intrinsic GTPase activity (Jaffé et al., 2012; Pandey et al., 2009). *COLD1* is predicted to contain a Ras GTPase-activating protein domain in the third cytoplasmic loop, and our biochemical data support this. Correspondingly, SNP2<sup>jap(A)</sup> versus ind(T/C) in fourth exon would cause an amino acid substitution in the third loop (Dong et al., 2007). Genetic complementation of *COLD1<sup>jap</sup>* instead of *COLD1<sup>ind</sup>* in *cold1-1* suggests that SNP2 functions in chilling tolerance (Figure 3B). The specific domain involved (i.e., the loop containing a predicted GTPase-activating protein domain) and its effects on GTPase activity, as well as  $\text{Ca}^{2+}$  signaling and electrophysiological response, are consistent with a *COLD1* biochemical function associated with G-protein signaling. We found that the substitution of Met<sup>187</sup>/Thr<sup>187</sup> for Lys<sup>187</sup> in *japonica* cultivars conferred stronger tolerance to chilling. Overexpression of *COLD1<sup>jap</sup>* also conferred enhanced tolerance. By contrast, the *COLD1<sup>ind</sup>* transgenic lines exhibited decreased tolerance, which could be explained by competition between *COLD1<sup>ind</sup>* and *COLD1<sup>jap</sup>* in interaction with RGA1 for regulation in  $[\text{Ca}^{2+}]_{\text{cyt}}$  level and GTPase activity (Figures 5 and 6).

Our genetic and biochemical analyses of *COLD1* revealed several similarities to mammalian cold receptors and plant heat sensors that lead us to hypothesize that *COLD1* is involved in sensing cold. (1) *COLD1* has broad tissue expression and is plasma- and ER-membrane localized, with nine predicted TM domains. (2) *COLD1* acts as a RGS to accelerate RGA1's GTPase activity and has phenotypic effects on chilling tolerance. (3) Cold-induced changes in  $\text{Ca}^{2+}$  influx and  $[\text{Ca}^{2+}]_{\text{cyt}}$  are mediated by *COLD1*. (4) Interaction between *COLD1* and RGA1 is required for the cold-induced specific electrophysiological response. (5) Differences in chilling tolerance are observed in *cold1-1*, in transgenic lines harboring various alleles from *japonica* and *indica*, and in the *RGA1* mutant, *d1*.

Cold temperature may be sensed through direct alteration of a sensor's structure and membrane fluidity to trigger cations influx for signaling. Notably, changes on  $\text{Ca}^{2+}$  signal involve both the resting level in the cytoplasm and the temporal elevation. The *cold1-1* showed lower resting levels of  $\text{Ca}^{2+}$ , which was genetically rescued by *COLD1<sup>jap</sup>* (Figure 6). This finding may hint that *COLD1* itself possibly represents a potential calcium permeable channel or a subunit of such a channel. Consequently, changes of this channel function would affect resting  $[\text{Ca}^{2+}]_{\text{cyt}}$ , which would influence the amplitudes of  $\text{Ca}^{2+}$  signals. The potential function of *COLD1* as a cold sensor could be simply explained by the lack of a significant  $\text{Ca}^{2+}$  gradient in *cold1-1* plants and *COLD1<sup>ind</sup>*-OE lines in  $\text{Ca}^{2+}$  resting levels that does not allow the formation of an appropriate  $\text{Ca}^{2+}$  signal. Therefore, it is appealing to speculate that *COLD1* is involved in sensing cold and that changes in



COLD1 protein structure and membrane fluidity in response to cold might initiate signaling through COLD1's physical interaction with RGA1, leading to  $\text{Ca}^{2+}$  influx into cytoplasm, which would then trigger downstream responses to chilling stress. Subsequently, accelerated GTPase activity of RGA1 by COLD1 might induce a regression shift on equilibrium between GDP- and GTP-bound states of RGA1 (Urano et al., 2012) (Figure S6).

The strong phenotype of plants with the *COLD1* QTL could result from tight functional interaction of COLD1 with important hormonal pathways. Consequently, an imbalance in COLD1 function likely affects multiple response pathways in this way aggravating the effects of its modulated temperature dependent functionality and thereby leading to significant decreased ability to re-assume growth after chilling stress. In this regard, COLD1 is functionally interconnected with the key gibberellin signaling component D1/RGA1 (Ueguchi-Tanaka et al., 2000) and brassinosteroid signaling, which are involved in regulation of plant height (Hu et al., 2013; Wang et al., 2006). Moreover, D1/RGA1 also affects TUD1, which mediates brassinosteroid signaling to regulate cell proliferation for plant growth and development (Hu et al., 2013; Wang et al., 2006). In addition D1/RGA1 is functionally dependent on SLR of GA signaling pathway for cell elongation (Ueguchi-Tanaka et al., 2000). In fact, our *cold1-1* significantly showed a decrease in plant height compared with wild-type, while plant height of the complemented lines of *cold1-1* with *COLD1* was recovered (Figure S3). Therefore, it is likely that COLD1 exhibits this strong impact on chilling tolerance via the RGA1 by disturbing multiple pathways, such as GA and/or BR signaling pathways (Hu et al., 2013; Wang et al., 2006).

We show here that a SNP of *COLD1* endows *japonica* rice with chilling tolerance, and that the mutation in the coding region of *COLD1* has been fixed in chilling-tolerant *japonica* cultivars. Our phylogenetic and population genetic analyses based on the large number of SNPs identified by resequencing 50 accessions of cultivated and wild rice (Huang et al., 2012; Xu et al., 2012) demonstrate that the chilling-tolerant allele originated from the Chinese *O. rufipogon* populations and was subject to strong human selection during *japonica* domestication, similar to the case of the *SD1* gene for *japonica* domestication (Asano et al., 2011). Therefore, genomic segments bearing agronomic traits can originate in one population and spread across all cultivars through artificial selection (He et al., 2011). Our findings are consistent with archaeological and genetic evidence that *japonica* rice was domesticated in China (Fuller et al., 2009; Huang et al., 2012; Londo et al., 2006; Xu et al., 2012). Importantly, our work demonstrates that the process of rice domestication was associated with fixation and extension of favored alleles or mutations that enhanced chilling tolerance for growth in regions with lower yearly temperatures. The *COLD1* allele and SNPs identified in this work have great potential for improving rice chilling tolerance via molecular breeding techniques.

## EXPERIMENTAL PROCEDURES

### Genetic Population and Plant Materials

*Oryza sativa* recombinant inbred lines (RIL) were developed by crossing *japonica* variety Nipponbare (NIP) and *indica* variety 93-11. The  $F_2$  generation from NIP  $\times$  93-11 was subjected to more than six rounds of self-pollination to

generate the RILs. For QTL genetic assay, the RILs were randomly selected. The near-isogenic lines were generated by backcrossing the NIP  $\times$  93-11 lines to 93-11 five times to generate BC<sub>5</sub>F<sub>2</sub>.

The T-DNA insertion mutant *cold1-1* was obtained from Dr G. An. *O. sativa* ssp. *japonica* cv. ZH10/11 and DJ were used for transformation to create the transgenic lines (Jeong et al., 2002). Mutant *cold1-1* was transformed with *COLD1* for a genetic complementation. The primers used for PCR are listed in Table S4.

### Chilling Treatment

To test chilling tolerance, the seedlings were treated at 2°C–4°C for various times based on the genetic background. Subsequently, they were moved to a temperature-controlled greenhouse with 28°C–30°C/25°C day/night cycles for recovery. After 3–7 days, the survival rate was determined as the percentage of the total seedlings that were alive (Ma et al., 2009).

### SNP Identification, Phylogenetic Analysis, Genetic Diversity, and Neutrality Tests

Full-length *COLD1* gene was sequenced using the tiling format. The primer sequences are listed in Table S4. The gene sequences from 127 samples were aligned using MEGA 5.0 software. A phylogenetic tree was constructed using the neighbor-joining method in MEGA5 (Tamura et al., 2011).

Estimates of nucleotide diversity and population genetic analyses were performed for each group using DnaSP 5.1 (Librado and Rozas, 2009). Tajima's D (Tajima, 1989) and maximum likelihood Hudson-Kreitman-Aguade (MLHKA) (Wright and Charlesworth, 2004) tests were used to examine the departure of *COLD1* polymorphisms from neutrality with a set of known neutral genes, namely, *Adh1*, *GBSSII*, *Ks1*, *Lhs1*, *Os0053*, *SSI1*, and *TFIIA $\gamma$ -1* (Zhu et al., 2007), as controls. The genome-wide controls with 400-kb regions around *COLD1* in 43 accessions were used for interpret the Tajima's statistics. The coalescent simulation analysis was carried out according to Wu et al. (2013). Details are in Supplemental Information.

### Subcellular Localization of COLD1

GFP was fused to COLD1 either at the N or C terminus. Its colocalization assays with marker proteins were carried out in protoplast (*Arabidopsis*, or Tobacco) cells as described previously (Lee et al., 2009). The transformed protoplast cells were examined by a confocal microscopy. See details in Supplemental Information.

### Coimmunoprecipitation Assay

Briefly, the recombinant plasmids were co-transformed into tobacco leaves according to Liu et al. (2007). The extracts were incubated with anti-FLAG M2 affinity gel (Sigma) or anti-GFP antibody at 4°C overnight. The antigen-antibody complex was collected. Then the sample was separated on SDS/PAGE gels for immunoblots. See details in Supplemental Information.

### Bimolecular Fluorescence Complementation

BiFC experiments and gene transformation were performed as described previously (Staglar et al., 1998; Waadt et al., 2008; Wang et al., 2009). The vectors were from Dr. J. Kudla. See details in Supplemental Information.

### Expression and Purification in *Spodoptera frugiperda*

Protein expression and purification of COLD1 in the cells of *Spodoptera frugiperda* (Sf9) were performed as previously described (Wu et al., 2010). Affinity chromatography was used in protein purification. See details in Supplemental Information.

### GTPase Activity Assay

The GTPase activity of RGA1 was monitored with the Enzcheck Phosphate Assay Kit as described previously (Dong et al., 2007). The amount of the tested protein (RGA1/COLD1 = 10/1  $\mu\text{g}$ ) was measured and confirmed in immunoblots using the FLAG antibodies. Amounts loaded were 1/0.1  $\mu\text{g}$  (RGA1/COLD1) for the blot. Details are in Supplemental Information.

### Electrophysiological Assay

For electrophysiological analysis, complementary RNA was prepared using the RNA Capping Kit (Stratagene). *Xenopus* oocytes were injected with

cRNA for *COLD1* and *RGA1*, mixed, and used for voltage-clamp experiments. Details are in [Supplemental Information](#).

### Extracellular $\text{Ca}^{2+}$ Flux and $[\text{Ca}^{2+}]_{\text{cyt}}$ Monitoring

The roots of 3-day-old seedlings were used to monitor  $\text{Ca}^{2+}$  flux with scanning ion-selective electrode technique (SIET) (Ludewig et al., 2003). The solution of 25°C was replaced with that of 0°C for the cold treatment.  $[\text{Ca}^{2+}]_{\text{cyt}}$  in callus was monitored by the cytosolic aequorin method (Saidi et al., 2009). The remaining aequorin was discharged by 1 M  $\text{CaCl}_2$  and 10% ethanol. Calibration of cytosolic  $\text{Ca}^{2+}$  concentration was according to Knight et al. (1996).

For monitoring  $\text{Ca}^{2+}$  elevation using Yellow Cameleon (YC3.6), whole plants were infected rice (GV3101) containing NES-YC3.6. Roots were used to monitor  $[\text{Ca}^{2+}]_{\text{cyt}}$  according to the method described by Krebs et al. (2012). Details are in [Supplemental Information](#).

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.046>.

### AUTHOR CONTRIBUTIONS

Y.M. performed experiments on phenotypic and biochemical assays. X.D. worked on the transgenic lines. Y.X. designed the experiments, analyzed data, and prepared the manuscript. W.L. performed genetic and evolution experiments. In Q.Q.'s lab, Q.Q. and D.Z. created the genetic population. In S.G.'s lab, S.G. and X.Z. performed analysis of molecular evolution. In W.W.'s lab, W.W. and X.X. performed SNP analysis. In L.L.'s lab, L.L. and Y.P. performed the electrical physiological assay. In J.J.'s lab, J.J. and X.L. performed the protein localization. H.L., D.Z., J.X., X.G., and S.X. performed some experiments on calcium and localization. H.Z. joined in calcium analysis. Y.N. performed gene transformation. K.C. designed all experiments, analyzed data, and wrote the manuscript.

### ACKNOWLEDGMENTS

The authors appreciate Dr. Qiang Zhao and Dr. B. Wu for their friendly help with purification of the membrane proteins. We are grateful to Dr. Joerg Kudla for his helpful comments and the kind gifts of constructs. We thank Dr. Gynheung An for the mutant line, and Ms. Rongxi Jiang, and Ms. Yuan Zhao, as well as Mr. Bo Wang for assistance with either gene transformation or molecular assay. We thank Dr. Yongbiao Xue, Dr. Bin Han, and Dr. Zhiyong Wang for their useful comments and help on the project. This work was supported by the Major State Basic Research Program of China (2013CBA01403) and the National Nature Science Foundation (30821007), as well as CAS grants (XDA08010205) to K.C.

Received: November 15, 2014

Revised: December 31, 2014

Accepted: January 14, 2015

Published: February 26, 2015

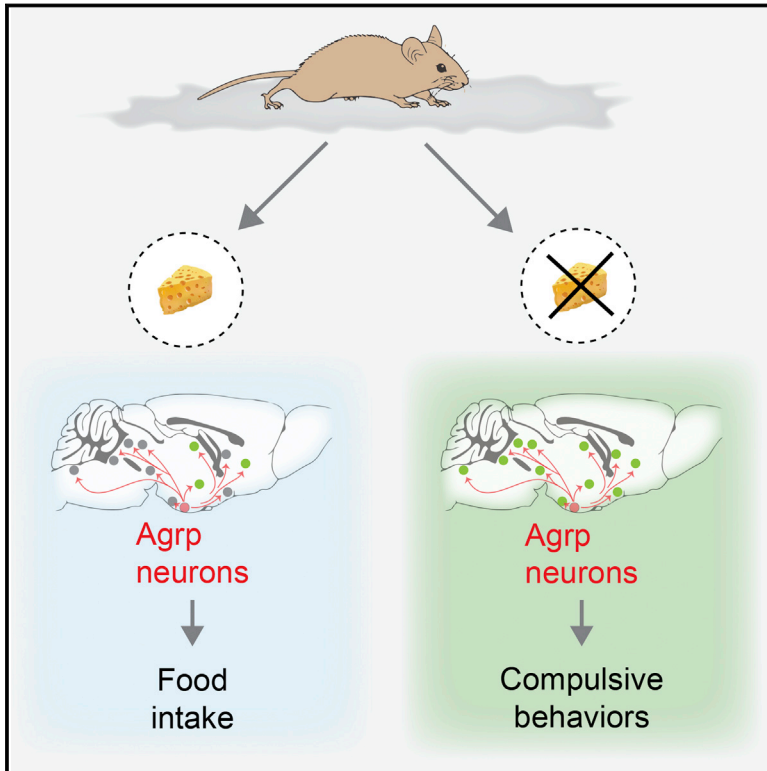
### REFERENCES

- Asano, K., Yamasaki, M., Takuno, S., Miura, K., Katagiri, S., Ito, T., Doi, K., Wu, J., Ebana, K., Matsumoto, T., et al. (2011). Artificial selection for a green revolution gene during japonica rice domestication. *Proc. Natl. Acad. Sci. USA* 108, 11034–11039.
- Batistic, O., Sorek, N., Schültke, S., Yalovsky, S., and Kudla, J. (2008). Dual fatty acyl modification determines the localization and plasma membrane targeting of CBL/CIPK  $\text{Ca}^{2+}$  signaling complexes in Arabidopsis. *Plant Cell* 20, 1346–1362.
- Chen, J.G., Willard, F.S., Huang, J., Liang, J., Chasse, S.A., Jones, A.M., and Siderovski, D.P. (2003). A seven-transmembrane RGS protein that modulates plant cell proliferation. *Science* 301, 1728–1731.
- Dai, X., Xu, Y., Ma, Q., Xu, W., Wang, T., Xue, Y., and Chong, K. (2007). Overexpression of an R1R2R3 MYB gene, *OsMYB3R-2*, increases tolerance to freezing, drought, and salt stress in transgenic *Arabidopsis*. *Plant Physiol.* 143, 1739–1751.
- Dong, G., Medkova, M., Novick, P., and Reinisch, K.M. (2007). A catalytic coiled coil: structural insights into the activation of the Rab GTPase Sec4p by Sec2p. *Mol. Cell* 25, 455–462.
- Finka, A., Cuendet, A.F., Maathuis, F.J., Saidi, Y., and Goloubinoff, P. (2012). Plasma membrane cyclic nucleotide gated calcium channels control land plant thermal sensing and acquired thermotolerance. *Plant Cell* 24, 3333–3348.
- Fujino, K., Sekiguchi, H., Matsuda, Y., Sugimoto, K., Ono, K., and Yano, M. (2008). Molecular identification of a major quantitative trait locus, *qLTG3-1*, controlling low-temperature germinability in rice. *Proc. Natl. Acad. Sci. USA* 105, 12623–12628.
- Fuller, D.Q., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L.A., and Sun, G.P. (2009). The domestication process and domestication rate in rice: spikelet bases from the Lower Yangtze. *Science* 323, 1607–1610.
- He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X., Greenberg, A.J., Hudson, R.R., Wu, C.I., and Shi, S. (2011). Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* 7, e1002100.
- Hu, X., Qian, Q., Xu, T., Zhang, Y., Dong, G., Gao, T., Xie, Q., and Xue, Y. (2013). The U-box E3 ubiquitin ligase TUD1 functions with a heterotrimeric G  $\alpha$  subunit to regulate Brassinosteroid-mediated growth in rice. *PLoS Genet.* 9, e1003391.
- Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501.
- Jaffé, F.W., Freschet, G.E., Valdes, B.M., Runions, J., Terry, M.J., and Williams, L.E. (2012). G protein-coupled receptor-type G proteins are required for light-dependent seedling growth and fertility in *Arabidopsis*. *Plant Cell* 24, 3649–3668.
- Jeong, D.H., An, S., Kang, H.G., Moon, S., Han, J.J., Park, S., Lee, H.S., An, K., and An, G. (2002). T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol.* 130, 1636–1644.
- Johnston, C.A., Taylor, J.P., Gao, Y., Kimple, A.J., Grigston, J.C., Chen, J.G., Siderovski, D.P., Jones, A.M., and Willard, F.S. (2007). GTPase acceleration as the rate-limiting step in Arabidopsis G protein-coupled sugar signaling. *Proc. Natl. Acad. Sci. USA* 104, 17317–17322.
- Knight, H., Trewavas, A.J., and Knight, M.R. (1996). Cold calcium signaling in Arabidopsis involves two cellular pools and a change in calcium signature after acclimation. *Plant Cell* 8, 489–503.
- Koseki, M., Kitazawa, N., Yonebayashi, S., Maehara, Y., Wang, Z.X., and Minobe, Y. (2010). Identification and fine mapping of a major quantitative trait locus originating from wild rice, controlling cold tolerance at the seedling stage. *Mol. Genet. Genomics* 284, 45–54.
- Kovach, M.J., Sweeney, M.T., and McCouch, S.R. (2007). New insights into the history of rice domestication. *Trends Genet.* 23, 578–587.
- Krebs, M., Held, K., Binder, A., Hashimoto, K., Den Herder, G., Parniske, M., Kudla, J., and Schumacher, K. (2012). FRET-based genetically encoded sensors allow high-resolution live cell imaging of  $\text{Ca}^{2+}$  dynamics. *Plant J.* 69, 181–192.
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Lee, H.K., Cho, S.K., Son, O., Xu, Z., Hwang, I., and Kim, W.T. (2009). Drought stress-induced Rma1H1, a RING membrane-anchor E3 ubiquitin ligase homolog, regulates aquaporin levels via ubiquitination in transgenic Arabidopsis plants. *Plant Cell* 21, 622–641.
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Liu, X., Yue, Y., Li, B., Nie, Y., Li, W., Wu, W.H., and Ma, L. (2007). A G protein-coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid. *Science* 315, 1712–1716.

- Londo, J.P., Chiang, Y.C., Hung, K.H., Chiang, T.Y., and Schaal, B.A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. USA* 103, 9578–9583.
- Ludewig, U., Wilken, S., Wu, B., Jost, W., Obrdlík, P., El Bakkoury, M., Marini, A.M., André, B., Hamacher, T., Boles, E., et al. (2003). Homo- and hetero-oligomerization of ammonium transporter-1 NH<sub>4</sub><sup>+</sup> uniporters. *J. Biol. Chem.* 278, 45603–45610.
- Ma, Q., Dai, X., Xu, Y., Guo, J., Liu, Y., Chen, N., Xiao, J., Zhang, D., Xu, Z., Zhang, X., and Chong, K. (2009). Enhanced tolerance to chilling stress in *Os-MYB3R-2* transgenic rice is mediated by alteration in cell cycle and ectopic expression of stress genes. *Plant Physiol.* 150, 244–256.
- Pandey, S., Nelson, D.C., and Assmann, S.M. (2009). Two novel GPCR-type G proteins are abscisic acid receptors in *Arabidopsis*. *Cell* 136, 136–148.
- Saidi, Y., Finka, A., Muriset, M., Bromberg, Z., Weiss, Y.G., Maathuis, F.J., and Goloubinoff, P. (2009). The heat shock response in moss plants is regulated by specific calcium-permeable channels in the plasma membrane. *Plant Cell* 21, 2829–2843.
- Saito, K., Miura, K., Nagano, K., Hayano-Saito, Y., Araki, H., and Kato, A. (2001). Identification of two closely linked quantitative trait loci for cold tolerance on chromosome 4 of rice and their association with anther length. *Theor. Appl. Genet.* 103, 862–868.
- Saito, K., Hayano-Saito, Y., Kuroki, M., and Sato, Y. (2010). Map-based cloning of the rice cold tolerance gene *Ctb1*. *Plant Sci.* 179, 97–102.
- Sang, T., and Ge, S. (2007). Genetics and phylogenetics of rice domestication. *Curr. Opin. Genet. Dev.* 17, 533–538.
- Sasaki, T., and Burr, B. (2000). International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. *Curr. Opin. Plant Biol.* 3, 138–141.
- Shabala, S., and Newman, I. (2000). Salinity effects on the activity of plasma membrane H<sup>+</sup> and Ca<sup>2+</sup> transporters in bean leaf mesophyll: Masking role of the cell wall. *Ann. Bot. (Lond.)* 85, 681–686.
- Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc. Natl. Acad. Sci. USA* 95, 5187–5192.
- Steinhorst, L., and Kudla, J. (2013). Calcium and reactive oxygen species rule the waves of signaling. *Plant Physiol.* 163, 471–485.
- Swarbreck, S.M., Colaço, R., and Davies, J.M. (2013). Plant calcium-permeable channels. *Plant Physiol.* 163, 514–522.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Ueguchi-Tanaka, M., Fujisawa, Y., Kobayashi, M., Ashikari, M., Iwasaki, Y., Kitano, H., and Matsuoka, M. (2000). Rice dwarf mutant d1, which is defective in the alpha subunit of the heterotrimeric G protein, affects gibberellin signal transduction. *Proc. Natl. Acad. Sci. USA* 97, 11638–11643.
- Urano, D., Phan, N., Jones, J.C., Yang, J., Huang, J., Grigston, J., Taylor, J.P., and Jones, A.M. (2012). Endocytosis of the seven-transmembrane RGS1 protein activates G-protein-coupled signalling in *Arabidopsis*. *Nat. Cell Biol.* 14, 1079–1088.
- Urano, D., Chen, J.G., Botella, J.R., and Jones, A.M. (2013). Heterotrimeric G protein signalling in the plant kingdom. *Open Biol.* 3, 120186.
- Waadt, R., Schmidt, L.K., Lohse, M., Hashimoto, K., Bock, R., and Kudla, J. (2008). Multicolor bimolecular fluorescence complementation reveals simultaneous formation of alternative CBL/CIPK complexes in planta. *Plant J.* 56, 505–516.
- Wang, L., and Chong, K. (2010). Signaling and communication in plants. In *Integrated G Proteins Signaling in Plants*, S. Yalovsky, F. Baluska, and A. Jones, eds. (Springer), pp. 1–25.
- Wang, L., Xu, Y.Y., Ma, Q.B., Li, D., Xu, Z.H., and Chong, K. (2006). Heterotrimeric G protein alpha subunit is involved in rice brassinosteroid response. *Cell Res.* 16, 916–922.
- Wang, L., Wang, Z., Xu, Y., Joo, S.H., Kim, S.K., Xue, Z., Xu, Z., Wang, Z., and Chong, K. (2009). OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice. *Plant J.* 57, 498–510.
- Wright, S.I., and Charlesworth, B. (2004). The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168, 1071–1076.
- Wu, B., Chien, E.Y., Mol, C.D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F.C., et al. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071.
- Wu, W., Zheng, X.M., Lu, G., Zhong, Z., Gao, H., Chen, L., Wu, C., Wang, H.J., Wang, Q., Zhou, K., et al. (2013). Association of functional nucleotide polymorphisms at DTH2 with the northward expansion of rice cultivation in Asia. *Proc. Natl. Acad. Sci. USA* 110, 2775–2780.
- Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., et al. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111.
- Yamauchi, S., Fusada, N., Hayashi, H., Utsumi, T., Uozumi, N., Endo, Y., and Tozawa, Y. (2010). The consensus motif for N-myristoylation of plant proteins in a wheat germ cell-free translation system. *FEBS J.* 277, 3596–3607.
- Zhu, Q., Zheng, X., Luo, J., Gaut, B.S., and Ge, S. (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* 24, 875–888.

# Hypothalamic Agrp Neurons Drive Stereotypic Behaviors beyond Feeding

## Graphical Abstract



## Authors

Marcelo O. Dietrich, Marcelo R. Zimmer, Jeremy Bober, Tamas L. Horvath

## Correspondence

marcelo.dietrich@yale.edu

## In Brief

Activity of Agrp neurons in the hypothalamus is known to stimulate feeding behavior, but activating them in the absence of food reveals that they also promote a set of stereotyped compulsive behaviors in mice.

## Highlights

- Agrp neuron activation leads to foraging and displacement behaviors
- Agrp neurons promote stereotypic behavioral responses
- Activation of Agrp neurons decreases anxiety levels
- $Y_5R$  signaling is necessary for Agrp neuron-mediated stereotypic behaviors





# Hypothalamic Agrp Neurons Drive Stereotypic Behaviors beyond Feeding

Marcelo O. Dietrich,<sup>1,2,3,\*</sup> Marcelo R. Zimmer,<sup>1,3</sup> Jeremy Bober,<sup>1</sup> and Tamas L. Horvath<sup>1,2,4</sup>

<sup>1</sup>Program in Integrative Cell Signaling and Neurobiology of Metabolism, Section of Comparative Medicine, Yale University School of Medicine, New Haven, CT 06520, USA

<sup>2</sup>Department of Neurobiology, Yale University School of Medicine, New Haven, CT 06520, USA

<sup>3</sup>Graduate Program in Biochemistry, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 90035, Brazil

<sup>4</sup>Kavli Institute for Neuroscience at Yale University, New Haven, CT 06520, USA

\*Correspondence: [marcelo.dietrich@yale.edu](mailto:marcelo.dietrich@yale.edu)

<http://dx.doi.org/10.1016/j.cell.2015.02.024>

## SUMMARY

The nervous system evolved to coordinate flexible goal-directed behaviors by integrating interoceptive and sensory information. Hypothalamic Agrp neurons are known to be crucial for feeding behavior. Here, however, we show that these neurons also orchestrate other complex behaviors in adult mice. Activation of Agrp neurons in the absence of food triggers foraging and repetitive behaviors, which are reverted by food consumption. These stereotypic behaviors that are triggered by Agrp neurons are coupled with decreased anxiety. NPY<sub>5</sub> receptor signaling is necessary to mediate the repetitive behaviors after Agrp neuron activation while having minor effects on feeding. Thus, we have unmasked a functional role for Agrp neurons in controlling repetitive behaviors mediated, at least in part, by neuro-peptidergic signaling. The findings reveal a new set of behaviors coupled to the energy homeostasis circuit and suggest potential therapeutic avenues for diseases with stereotypic behaviors.

## INTRODUCTION

Neural circuits are responsible for organizing and regulating flexible goal-oriented behaviors by integrating sensory and interoceptive information. The observation that mice can perform complex dynamic computations similar to humans (Kheifets and Gallistel, 2012) supports the view that brain mechanisms involved in complex goal-oriented behaviors rely on phylogenetically primitive neural circuits.

Homeostatic functions—for example, food intake—are adaptive responses that allow successful survival of the individual in the environment. The hypothalamus is an ancient brain region present in all vertebrates that is critical for the regulation of homeostatic functions, including energy balance, sexual behavior, sleep, and thirst. For more than 20 years, hypothalamic neurons that produce NPY, Agrp, and GABA have been thought to be involved in the promotion of hunger (Hahn et al., 1998; Horvath et al., 1992; Horvath et al., 1997). Neuropeptide injections in

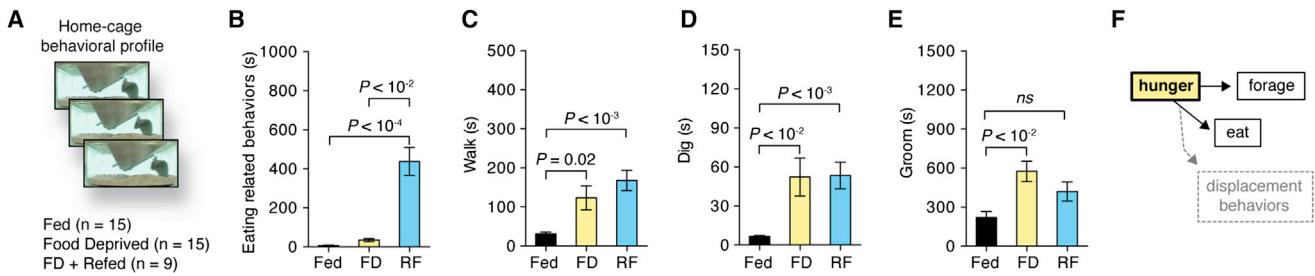
the brain elicit robust increases in food intake (Clark et al., 1984; Ollmann et al., 1997; Rossi et al., 1998; Stanley et al., 1986), and food deprivation increases the activity of these neurons (Hahn et al., 1998; Liu et al., 2012; Takahashi and Cone, 2005; Yang et al., 2011). Acute (Gropp et al., 2005; Luquet et al., 2005), but not chronic (Xu et al., 2005), ablation of Agrp neurons leads to cessation of feeding and, ultimately, death (Luquet et al., 2005). Conversely, acute activation of these neurons induces robust feeding (Aponte et al., 2011; Krashes et al., 2011). The neural circuits involved in the regulation of hunger by Agrp neurons seem to involve several brain nuclei (Atasoy et al., 2012; Betley et al., 2013; Wu et al., 2012). Agrp neurons have a broad projection field (Broberger et al., 1998) with important developmental characteristics as well (Dietrich et al., 2012; Grove et al., 2001). It is, therefore, intuitive to postulate that Agrp neurons orchestrate complex behavioral and physiological changes that encompass hunger rather than just food intake. This hypothesis gains momentum when neuropsychiatric conditions with strong homeostatic components are considered (e.g., anorexia nervosa). For instance, anorexia nervosa is a state of severe negative energy balance, in which brain circuits controlling feeding may be involved in the development of cognitive impairments of this disorder.

Here, we tested these assumptions by performing analysis of mouse behavior under conditions of Agrp neuron activation. Our results uncover a fundamental role for Agrp neuron activation in promoting repetitive/stereotypic behaviors in mice, unmasking a previously unsuspected role for these hypothalamic neurons.

## RESULTS

### Hunger-Related Behaviors

We first determined the effects of food deprivation, a physiological state of elevated Agrp neuronal activity (Hahn et al., 1998; Takahashi and Cone, 2005), on behavior. We used software-assisted characterization of mouse home-cage behaviors (Adamah-Biassi et al., 2013; Jhuang et al., 2010; Kyzar et al., 2012) to assess different aspects of the behavioral repertoire that occurs during hunger (Figure 1A). We studied fed, food deprived (FD), and food-deprived mice that were re-fed (RF). We divided our analysis into three large groups of behaviors: (1) consummatory responses represented by eating-related behaviors (e.g., time spent in the eating zone and chewing); (2) appetitive



**Figure 1. Home-Cage Behaviors in Food-Deprived and Re-Fed Mice**

(A) Mouse behaviors in the home cage of fed (black bars), food-deprived (yellow bars), and re-fed (blue bars) mice.

(B–E) Time spent in (B) eating-related behaviors, (C) walking, (D) digging, and (E) grooming.

(F) Behaviors elicited by hunger states.

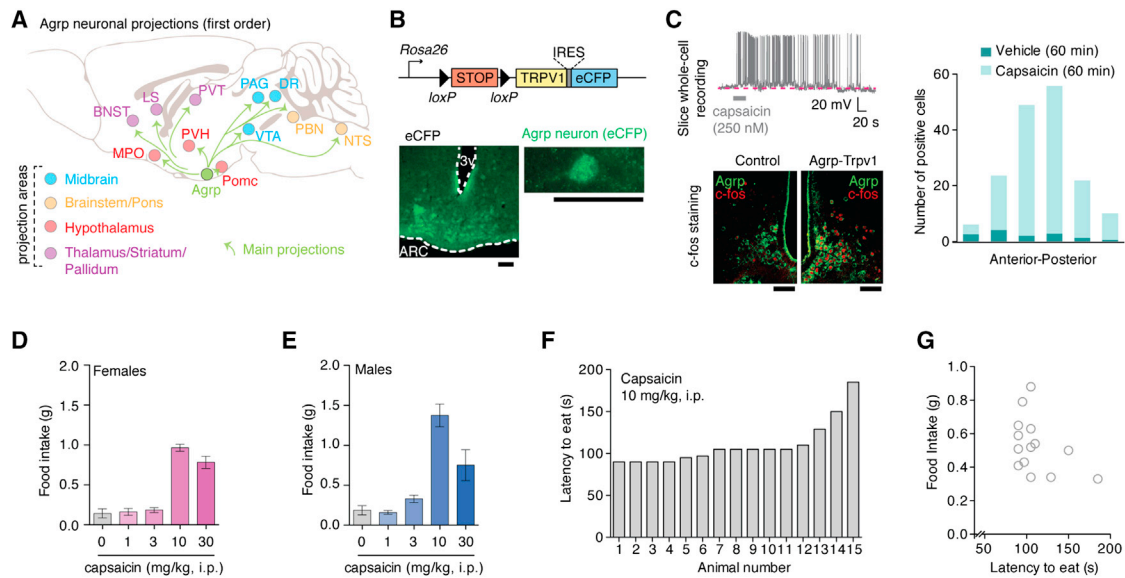
Error bars represent mean  $\pm$  SEM. *p* values represent Holm-Sidak's multiple comparisons test.

behaviors (forage-related behaviors, e.g., digging and walking); and (3) displacement behaviors (e.g., grooming). As expected, fed and FD animals did not engage in eating-related behaviors when food was not presented in the home cage, an effect promptly reverted in re-fed animals (Figure 1B). Food deprivation stimulated forage-related behaviors, an effect that persisted in the re-fed group (Figures 1C and 1D). Because our analyses lasted for 1 hr after the introduction of food to mice, our data indicate that the mechanisms involved in foraging behaviors during food deprivation are slowly switched off by satiety and not acutely by immediate presentation of food. Food deprivation also exacerbated grooming behavior (Figure 1E). In such conditions, grooming has been considered a displacement behavior (Barnett, 1956), a substitute of consummatory eating. Re-feeding acutely attenuated grooming (Figure 1E), reinforcing that displacement behaviors, such as grooming, manifest when animals lack the consummatory response. Thus, hunger promotes foraging (appetitive), eating (consummatory), and grooming (displacement) behaviors in mice (Figure 1F). Because the activation of *Agrp* neurons promotes hunger in sated mice (Aponte et al., 2011; Krashes et al., 2009; Krashes et al., 2011), we next asked what aspects of the behavior repertoire promoted by food deprivation may be induced by acute activation of the *Agrp* neurons.

### Acute *Agrp* Neuronal Activation

*Agrp* neurons have a broad projection field (Broberger et al., 1998), which extends to a wide range of subcortical areas (Figure 2A). This complex connectivity indicates that *Agrp* neurons have the capability to modulate a broad range of behaviors using multiple parallel circuits. In a previous study, we showed that *Agrp* neurons influence motivational states not related to feeding—for example, responses to cocaine (Dietrich et al., 2012). As an underlying mechanism, our data indicated that *Agrp* neurons have a developmental effect on dopamine cell function. These data reinforce the notion that animal models with altered *Agrp* neuronal activity during development are not suitable for the study of their acute role in the adult (Dietrich et al., 2012). Here, to examine the acute effects of *Agrp* neurons on adult animal behavior, we utilized animal models that allowed activation of *Agrp* neurons in a rapid, reliable, and reproducible manner.

Several techniques have been developed to acutely manipulate neuronal function *in vivo*. Optogenetics (Aponte et al., 2011) and chemical genetics using designer receptors exclusively activated by designer drugs (DREADDs) (Krashes et al., 2011) have been used to study the effects of *Agrp* neuron activity on the feeding behavior of adult mice. Optogenetics provide good time resolution with early onset of feeding behavior (Aponte et al., 2011); however, it requires the insertion of a light source deep into the brain, which adds a bias when analyzing complex behaviors. On the other hand, DREADD can be used to activate *Agrp* neurons by peripherally injecting receptor-ligand with robust induction of food intake (Krashes et al., 2011) but with more coarse kinetics (Rogan and Roth, 2011). We used transgenic mice that conditionally express *Trpv1* in *Cre*-expressing cells (Arenkiel et al., 2008; Güler et al., 2012) (*R26-LSL-Trpv1*; Figure 2B) to selectively introduce *Trpv1* in *Agrp* neurons. By backcrossing these mice (*R26-LSL-Trpv1*) to a *Trpv1* knockout background and then to *Agrp-Cre* mice, we generated animals that express *Trpv1* exclusively in the *Agrp* neurons (hereafter, *Agrp-Trpv1* mice; Figures 2B and S1). We performed a series of control experiments to confirm that expression of *Trpv1* was restricted to *Agrp* neurons in the arcuate nucleus and not in off-target cells (Figure S1 and Experimental Procedures). *Trpv1* is a cation channel that is activated by the exogenous agonist capsaicin (Caterina et al., 1997) in a rapid and reversible manner (Güler et al., 2012). Slice whole-cell recordings showed that capsaicin increased the firing rate of *Agrp* neurons (Figure 2C). The analysis of *c-fos* expression in *Agrp* neurons after capsaicin injection (i.p.) in *Agrp-Trpv1* mice revealed that most *Agrp* neurons throughout the arcuate nucleus were activated in these transgenic mice (Figure 2C). Capsaicin injection of *Agrp-Trpv1* mice led to increased food intake in both female (Figure 2D) and male mice (Figure 2E and Movie S1). Notably, the amount of food consumed by the activation of *Agrp* neurons in our studies was of similar magnitude as that observed when these cells were activated by optogenetics or DREADDs (Aponte et al., 2011; Krashes et al., 2011). The latency to eat in *Agrp-Trpv1* mice was faster (mean = 110.1 s [95% CI = 95.5–124.6], *n* = 15 mice) compared to these other techniques (Aponte et al., 2011; Krashes et al., 2011) (Figures 2F and 2G). Thus, this animal model enabled us to rapidly and reliably activate *Agrp* neurons by peripheral injection of capsaicin and explore their role on behaviors.



**Figure 2. Trpv1 Channels in AgRP Neurons Allow Acute Control of Neuronal Activity**

(A) Main projection from AgRP neurons.  
 (B) Reporter Trpv1 mice and CFP staining in the arcuate nucleus of AgRP-Trpv1 mice.  
 (C) Whole-cell recording of an AgRP-Trpv1 neuron and c-fos staining in AgRP-Trpv1-HA reporter mice 60 min after capsaicin injection (10 mg/kg, i.p.).  
 (D and E) Food intake in (D) female AgRP-Trpv1 and in (E) male mice.  
 (F) Latency to eat in female AgRP-Trpv1 mice.  
 (G) Correlation between latency and food intake.

Error bars represent mean  $\pm$  SEM. Scale bars, 50  $\mu$ m. See also Figure S1 and Movie S1.

### Repertoire of Home-Cage Behaviors

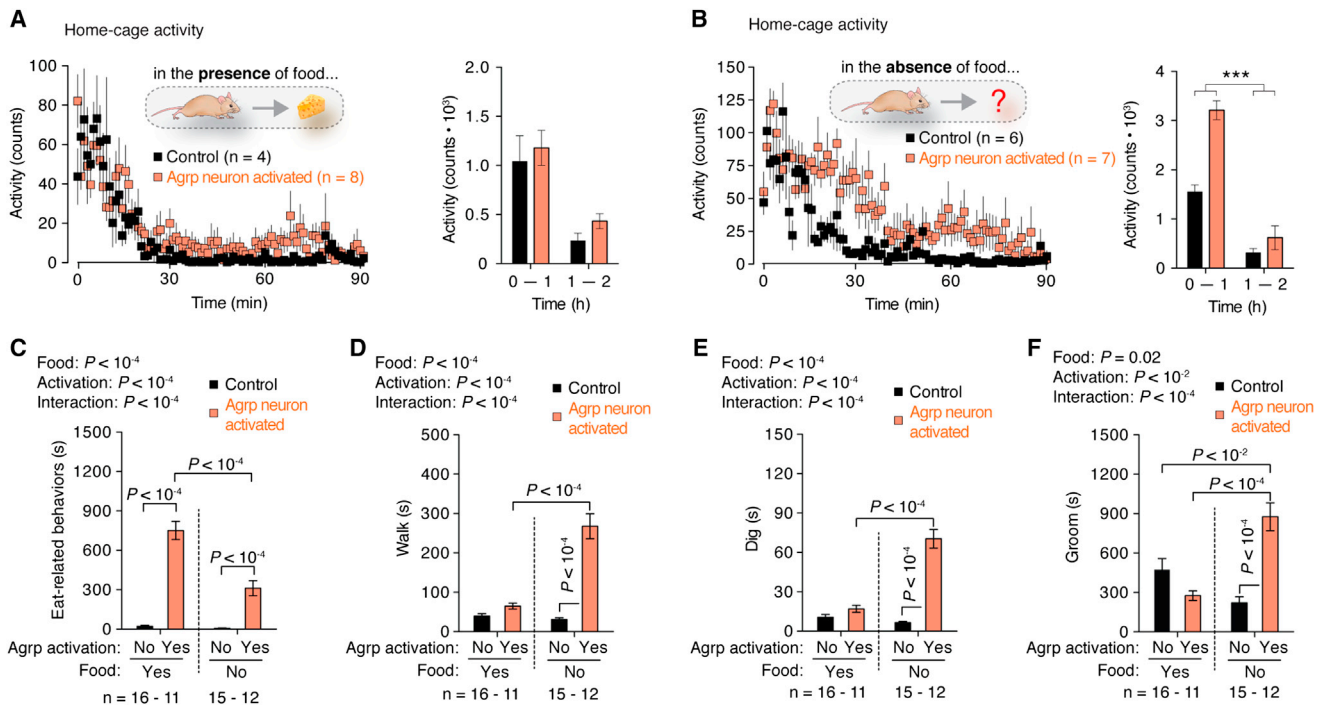
To screen for broad changes in behavior after AgRP neuron activation, we investigated changes in home-cage behaviors in the presence or absence of food in sated mice (Figure 3). In the presence of food, activation of AgRP neurons did not statistically change ambulatory activity (Figure 3A), while it evoked feeding in all AgRP-Trpv1 mice tested. Conversely, when food was removed, AgRP neuron activation increased activity levels (Figure 3B). To dissect these behavioral changes, we characterized mouse behaviors in their home cages upon activation of the AgRP neurons, similarly to what we did in FD mice (Figure 1). These experiments were performed in sated mice provided with food or with an empty food container. In all AgRP-Trpv1 mice tested in this paradigm, when food was present in their home cage, injection of capsaicin evoked robust food intake (data not shown). As expected, consummatory aspects of feeding, as measured by eating-related behaviors, were greatly enhanced by AgRP neuron activation (Figure 3C). Interestingly, activation of AgRP neurons in sated mice in the absence of food also led to increases in eating-related behaviors (e.g., interaction with the empty food container and chew bedding material; Figure 3C). The persistence of these behaviors indicates a degree of repetitiveness and stereotypy in the behavior repertoire of AgRP neuron activated animals in the absence of food.

Forage-related behaviors were increased in AgRP-neuron-activated mice in the absence of food, an effect that was almost completely reverted in the presence of food (Figures 3D and 3E). Grooming also increased after treatment of AgRP-Trpv1 mice with capsaicin in the absence of food but decreased when

animals were provided food (Figure 3F). Grooming is considered a displacement behavior to attenuate the appetitive response (forage) in the absence of the stimulus (food). When manifested in excess, grooming has also been related to obsessive-compulsive behaviors in mice (Ahmari et al., 2013; Burguière et al., 2013), similar to digging (Karvat and Kimchi, 2012). Thus, our findings indicate that, in addition to appetitive and consummatory aspects of hunger, the activation of AgRP neurons in AgRP-Trpv1 mice is sufficient to drive repetitive/stereotypic behaviors, an unsuspected role for these hypothalamic neurons. To corroborate these findings, we expressed hM3Dq in AgRP neurons by injecting AgRP-Cre mice with a recombinant AAV vector carrying a cre-dependent coding sequence (rAAV-FLEX-hM3Dq-mCherry). The activation of AgRP neurons by peripheral injection of the receptor ligand, clozapine-N-oxide (CNO, 0.3 mg/kg, i.p.), led to similar results as observed in AgRP-Trpv1 mice injected with capsaicin (Figure S2) but with a delayed response, consistent with the slow effect of hM3Dq in stimulating neuronal activity (Krashe et al., 2011; Rogan and Roth, 2011). Altogether, we conclude that activation of the AgRP neurons resembles many, but not all, aspects of food deprivation. Our findings place interoceptive regions of the mammalian brain, such as the arcuate nucleus of the hypothalamus, as crucial mediators of repetitive and stereotypic behaviors (Figures 3C and 3F). Thus, we set out to investigate these behavioral responses in greater detail.

### AgRP Neurons Trigger Repetitive Behaviors

To further evaluate the extent to which the activation of AgRP neurons can engage mice in repetitive behaviors, we tested



**Figure 3. Home-Cage Behavior Analysis of AgRP Neuronal Activated Mice**

(A) Activity in the home cage with food provided.

(B) Activity with no food provided.

(C) Eat-related behaviors.

(D) Time walking.

(E) Time digging.

(F) Time grooming.

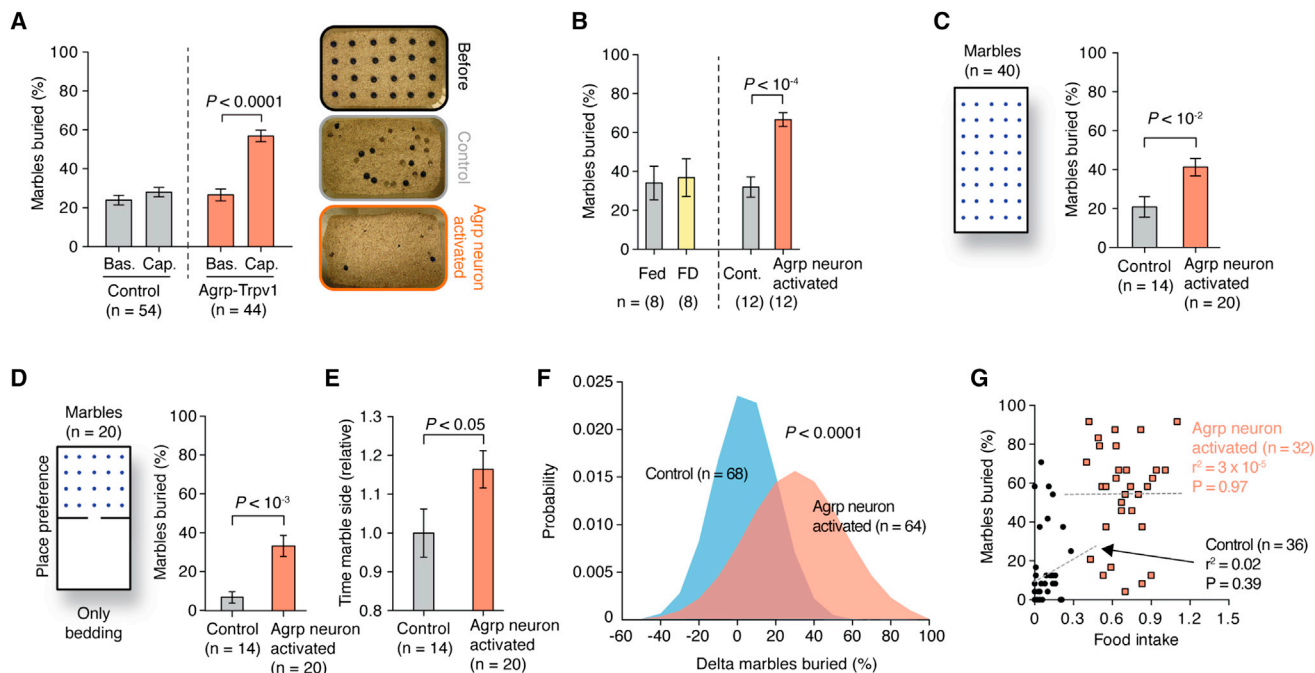
Symbols and bars represent mean  $\pm$  SEM. Statistical data derived from two-way ANOVA and Holm-Sidak's multiple comparisons test. See also Figure S2.

AgRP-Trpv1 mice in the marble-burying test (Deacon, 2006; Gyertyán, 1995; Witkin, 2008). The activation of AgRP neurons led to a robust increase in the number of marbles buried by males (Figure 4A and Movies S2 and S3) and females (data not shown), an effect that was, at least, in the same order of magnitude as mouse models of obsessive-compulsive disorders (Amodeo et al., 2012). Because food deprivation increases digging and grooming in the absence of food (Figure 1), which can also be considered repetitive behaviors (Ahmari et al., 2013; Burguière et al., 2013; Karvat and Kimchi, 2012), we tested food-deprived mice in parallel to AgRP-neuron-activated mice in the marble-burying test. We did not find statistical differences in the number of marbles buried after food deprivation (Figure 4B). To further test whether the increase in marble-burying behavior was due to repetitiveness, we performed a modified marble-burying test. We assessed mice in a larger cage with 40 marbles, which decreases the overall number of marbles buried and increases exploratory behavior. We found similar data in this modified version of the marble-burying test, with activation of AgRP neurons increasing the number of marbles buried (Figure 4C) while decreasing total activity during the test (control =  $42.96 \pm 3.12$  m [n = 14], AgRP-Trpv1 =  $32.27 \pm 2.05$  m [n = 20, mean  $\pm$  SEM];  $p = 0.004$ , two-tailed Mann-Whitney test), likely due to the extended time that mice spent burying marbles rather than

exploring the arena. To test whether chronic negative energy balance impacts AgRP neuron activation responses, we placed animals on a 20% calorie-restricted regimen for 4 weeks and then tested them. Similar to the ad libitum fed animals (Figure 4), the activation of AgRP neurons by capsaicin increased marble-burying behavior in calorie-restricted mice (Figure S3). These results, together with the data gained in sated mice, argue for the importance of AgRP neuronal activity rather than metabolic state per se as a controller of stereotypic behaviors.

To further investigate whether the increase in marble burying was due to a goal-oriented repetitive behavior (to bury marbles) (Gyertyán, 1995; Londei et al., 1998; Thomas et al., 2009), we performed a place preference test (Figure 4D). Marbles were distributed on only one side of the cage, and bedding was present on both sides. AgRP-Trpv1 mice that received capsaicin buried a much larger number of marbles (Figure 4D) and spent  $\sim 16\%$  more time on the marble side of the chamber (Figure 4E) than control mice. Notably, even with only half of the cage covered with marbles (Figure 4D), the number of marbles buried did not differ from the previous experiment (Figure 4C) in AgRP-Trpv1 mice injected with capsaicin (full cage =  $41.25 \pm 4.55$  [n = 20]; half cage =  $33.00 \pm 5.45$  [n = 20, mean  $\pm$  SEM];  $p = 0.183$ , two-tailed Mann-Whitney test) but decreased in the control group (full cage =  $20.89 \pm 5.25$  [n = 14]; half





**Figure 4. Repetitive Behaviors after Agrp Neuron Activation**

(A) Marbles buried after Agrp neuron activation.  
 (B) Marble buried in fed, food-deprived (FD), control, and Agrp-neuron-activated mice.  
 (C) Marble buried in the modified marble-burying test.  
 (D) Marble buried in the modified place-preference test.  
 (E) Time animals spent in the marble side relative to control animals.  
 (F) Normal distribution fitted to pooled experimental data (delta marbles buried [capsaicin injection – baseline]). p value was calculated using unpaired t test with Welch's correction.  
 (G) Linear regression analysis correlating marble-burying behavior and food intake. Each data point represents one mouse. Female mice were used in this study. Error bars represent mean  $\pm$  SEM, and p values were calculated using t test. See also Figure S3 and Movies S2 and S3.

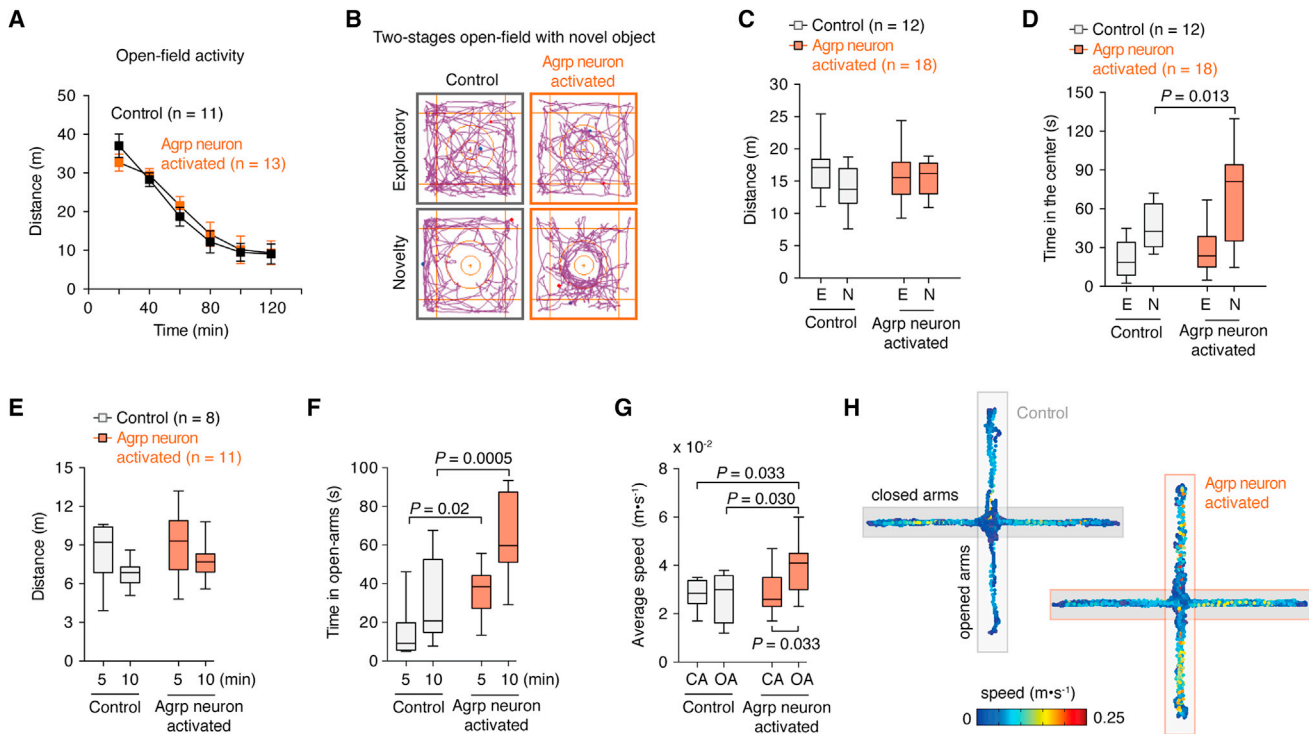
cage =  $6.78\% \pm 2.80\%$  [n = 14, mean  $\pm$  SEM]; p = 0.01, two-tailed Mann-Whitney test], indicating that the activation of Agrp neurons directs the animal's behavior toward repetitive, stereotypic responses when food is not available.

We hypothesized that, if Agrp neuron-mediated feeding and repetitive behaviors are a result of the same brain circuit, then these two behaviors should be correlated. We took advantage of the marble-burying behavior to test repetitive responses in mice. Frequency distribution histograms show a shift to the right in the number of marbles buried in Agrp-neuron-activated mice (Figure 4F), highlighting the idea that these behavioral changes are variable and affect differently subpopulations of mice. Linear regression analysis of individual responses did not show a correlation between marble-burying and feeding behaviors (Figure 4G), suggesting that the brain circuits that drive these behaviors by Agrp neurons are distinct and not completely overlapping.

#### Agrp Neuron Activation Decreases Anxiety

It is possible that changes in repetitive and stereotypic behaviors observed after Agrp neuron activation are due to increased anxiety. It is expected that treatments that increase anxiety levels will also increase repetitive/stereotypic responses in mice. Hunger is an unpleasant physiological state. Thus, it is possible that the promotion of hunger by activation of Agrp neurons generates

an anxiogenic state in mice that leads to repetitive behaviors, as described above. To test anxiety-related behaviors, we performed a series of tests. First, we placed mice in a novel open-field exploratory test following activation of Agrp neurons by capsaicin. We did not find significant changes in total activity (Figure 5A) or time that animals explored the center of the arena (data not shown). We then put mice in a two-stage open-field test, in which a novel object is added to the center of the arena to induce novelty exploration and anxiety (Dietrich et al., 2012). In this test, activation of Agrp neurons increased the time that animals spent exploring the object (Figures 5B and 5D), but not total activity (Figure 5C). This indicates a decrease in anxiety levels compared to control mice. Next, we assessed mice in the zero- and plus-maze apparatuses, in which anxiety-related behaviors inversely correlate with the time that animals spend in the open arms. In both tests, we did not observe significant changes in activity levels between groups (Figures 5E and S4 and Movies S4 and S5), but we found that the activation of Agrp neurons increased the time in the open arms (Figures 5F and S4). Intriguingly, Agrp-neuron-activated mice accelerated once in the open arms (Figures 5G-H), perhaps due to changes in risk assessment. This hypothesis needs further investigation. Overall, the data show that activation of Agrp neurons in mice leads to repetitive behaviors that are not due to increases in anxiety levels.



**Figure 5. Activation of AgRP Neurons Decreases Anxiety-Related Behaviors**

(A) Activity in the open field. Data points represent mean  $\pm$  SEM.

(B) Two-stage open-field test.

(C) Total distance traveled in the two-stage open-field test.

(D) Time spent in the center of the open field.

(E) Distance traveled by mice in the plus-maze test.

(F) Time animals spent in the open arms.

(G) Average speed of mice in the close arms (CA) and open arms (OA) of the apparatus.

(H) Representative tracking data.

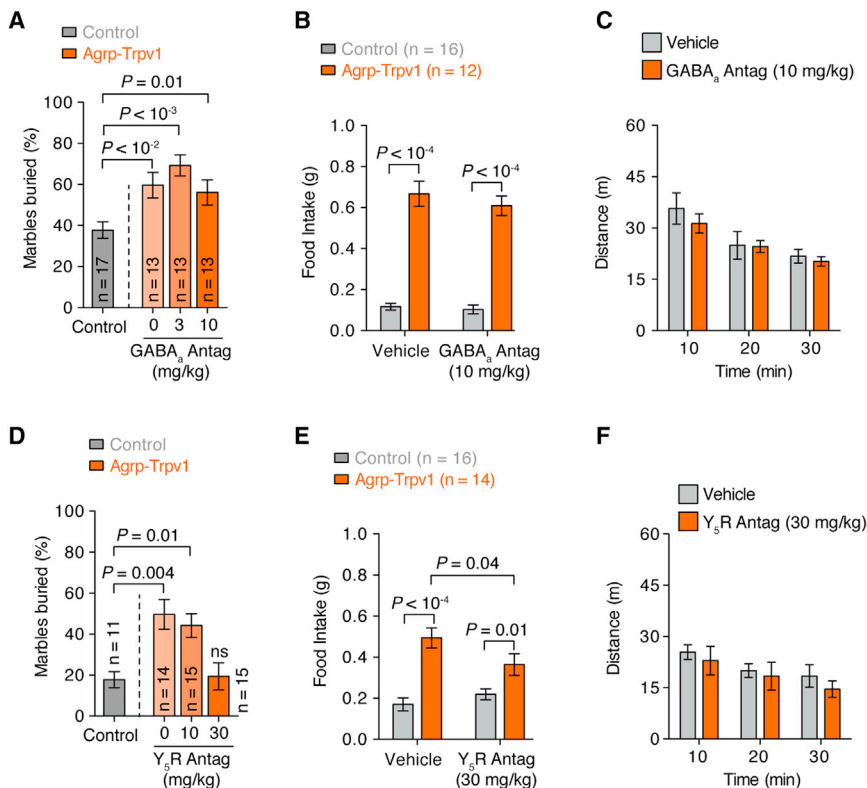
See also Figure S4 and Movies S4 and S5. Box and whiskers represent median  $\pm$  min/max values. p values were calculated using two-way ANOVA with repeated-measures followed by Holm-Sidak's multiple comparisons test.

Conversely, the activation of AgRP neurons is anxiolytic in several behavior tests.

### Alleviation of Behaviors by Y5 Receptor Antagonist

AgRP neurons have been shown to induce voracious food intake after acute activation due to NPY and GABA release (Aponte et al., 2011; Krashes et al., 2013). Because animal models in which GABA and/or NPY signaling is removed from AgRP neurons have developmental consequences (Atasoy et al., 2012; Dietrich et al., 2012), we examined whether pharmacological blockage of these signaling pathways would prevent repetitive behaviors after AgRP neuron activation. Systemic injection of a GABA<sub>A</sub> receptor antagonist was unable to reverse the induction of marble-burying behavior (Figure 6A) and food intake (Figure 6B) in AgRP-Trpv1 mice injected with capsaicin. NPY from the arcuate nucleus seems to signal mostly through NPY<sub>1</sub> and NPY<sub>5</sub> receptors, with overlapping expression and function (Atasoy et al., 2012; Gerald et al., 1996; Kanatani et al., 2000; Pedrazzini et al., 1998; Wolak et al., 2003). We have shown an anatomical link between the lateral hypothalamic orexin/hypocretin neurons and the arcuate nucleus NPY/AgRP cells (Horvath

et al., 1999). Neuropeptides released by orexin/hypocretin neurons promote feeding, an effect that we showed to be diminished by administration of a NPY<sub>5</sub> receptor antagonist (Dube et al., 2000). These previous observations together with the translatability of NPY<sub>5</sub> receptor antagonists (Erundu et al., 2006) led us to interrogate the role of NPY<sub>5</sub> receptor signaling in behavioral changes mediated by AgRP neuron activation. Systemic injection of a NPY<sub>5</sub> receptor antagonist before activation of AgRP neurons was sufficient to block the increase in marble-burying behavior (Figure 6D) while slightly decreasing food intake (Figure 6E). Neither GABA<sub>A</sub> receptor nor NPY<sub>5</sub> receptor antagonists altered locomotor activity in an open field at the maximum dose used in this study (Figures 6C and 6F). These results indicate that NPY<sub>5</sub> receptor signaling is necessary for the repetitive behaviors induced by the activation of AgRP neurons. To further evaluate the participation of NPY<sub>5</sub> receptor signaling in the behavior repertoire of mice after AgRP neuronal activation, we scrutinized mouse behavior in the home cage. We treated mice with the NPY<sub>5</sub> receptor blocker before activating AgRP neurons by capsaicin in AgRP-Trpv1 mice (Figure 7A). While activation of AgRP neurons increased eating-related (Figure 7B) and foraging-related



**Figure 6. Effects of GABA<sub>A</sub> or NPY<sub>5</sub> Receptors Blockade in Agrp-Neuron-Activated Mice**

(A) Effect of the GABA<sub>A</sub> receptor blocker, bicuculline, in the marble-burying test after activation of Agrp neurons.

(B) Effect of bicuculline on food intake.

(C) Effect of bicuculline on locomotor activity.

(D) Similar to A but using the NPY<sub>5</sub> receptor antagonist (CGP71683 hydrochloride).

(E) Similar to B using CGP71683.

(F) Similar to C using CGP71683.

Error bars represent mean ± SEM. p values were calculated using one-way ANOVA in A and D and two-way ANOVA with repeated-measures in B, C, E, and F followed by Holm-Sidak's multiple comparisons test.

behaviors (Figures 7C–7E), blockage of NPY<sub>5</sub> receptor signaling attenuated all of these behavioral responses with no effects in control mice (Figures 7B–7E). Remarkably, the effects of Agrp neuron activation on grooming were completely reverted by systemic injection of NPY<sub>5</sub> receptor blocker (Figures 7F–7H), similar to the effects reported in the marble-burying experiment (Figure 6D). Thus, we found that activation of Agrp neurons leads to repetitive behaviors, a behavioral phenotype that is completely reverted by NPY<sub>5</sub> receptor blockade. Notably, treatment of control mice with a NPY<sub>5</sub> receptor antagonist did not significantly alter baseline behaviors, but only behaviors driven by Agrp neuron activation. Because feeding response is not fully reverted by blocking NPY<sub>5</sub> receptor signaling (Figure 6E) and because repetitive and feeding responses are not correlated behaviors (Figure 4G), our data provide further support for the idea that different Agrp neuronal subpopulations promote food intake versus repetitive/stereotypic behaviors (Figure S5).

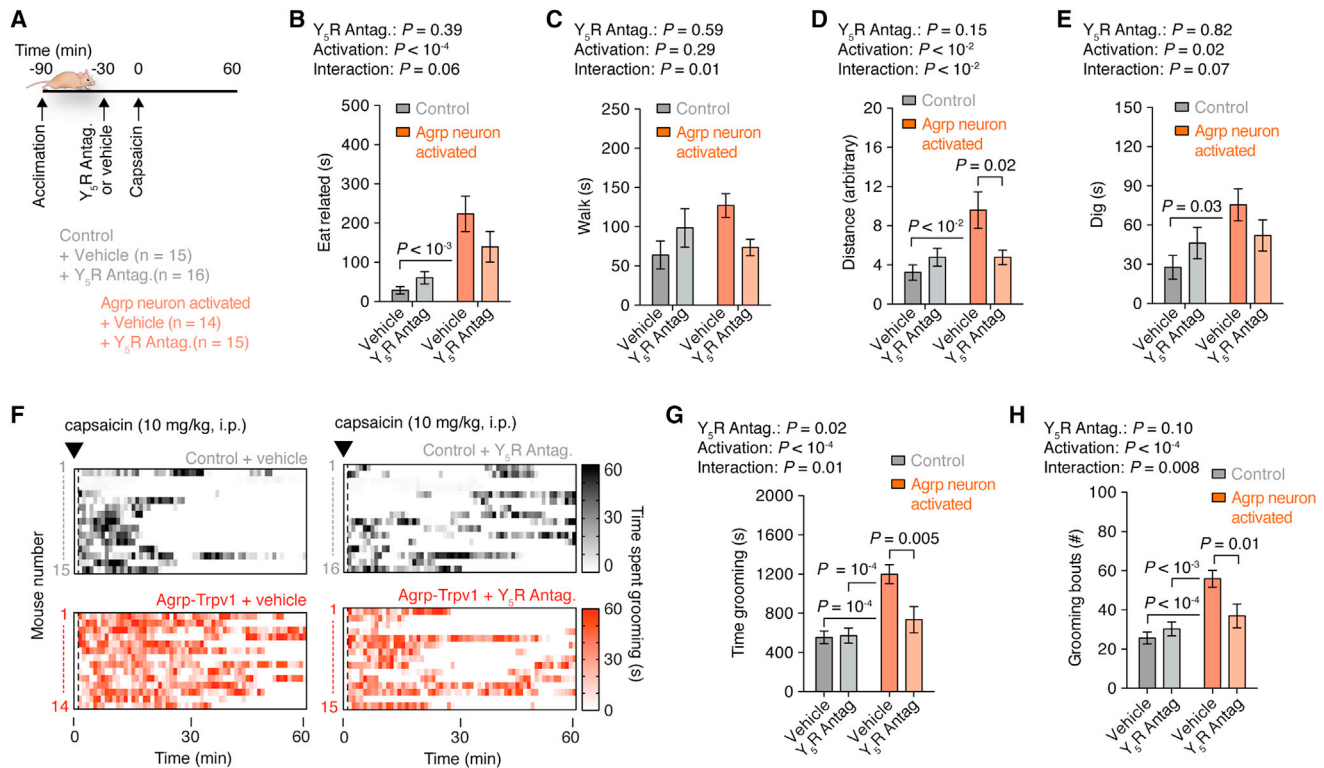
## DISCUSSION

The hypothalamus integrates hormonal and ascending neural inputs that bring information from the periphery (Chaudhri et al., 2006; Coll et al., 2007; Dietrich and Horvath, 2009; Lam et al., 2005). Our findings highlight the importance of Agrp neurons in mediating the effect of the peripheral environment on complex brain functions and behaviors. Our results identified the hypothalamic Agrp neurons as initiators of stereotypic behaviors in mice. These behaviors were triggered when the vast majority of Agrp neurons were simultaneously activated. Some aspects of the

stereotypic behaviors induced by chemical genetic activation of Agrp neurons were not seen in food-deprived animals or calorie-restricted mice. These observations suggest that different subpopulations of Agrp neurons subserve different functions, and it is likely that their activity patterns are not synchronized and are under differential input control. The fact that some behavioral shifts induced by Agrp neuronal activation can while others

cannot be suppressed by a NPY<sub>5</sub> receptor blocker further argue for the segregation of function of different subpopulations of Agrp cells. Thus, it is anticipated that an intricate and highly complex input organization and efferent connectivity of various subpopulations of Agrp neurons exists to support predictable and dynamic behavioral and autonomic adaptations to the changing environment (Figure S5).

Our results unmasked a previously unsuspected role for the hypothalamic hunger-promoting neurons in controlling repetitive, stereotypic behaviors in mice. Also, we showed that the activation of Agrp neurons decreases anxiety levels in several tests in mice. Because the hypothalamus is an evolutionarily conserved brain region, it is likely that these results are relevant to higher-order organisms, including humans. A recent report reinforces this view by providing evidence that mice are capable of estimating probabilities and calculating risks to make behavioral adjustments in dynamic environments analogous to humans (Kheifets and Gallistel, 2012). This supports the argument that brain mechanisms involved in complex behaviors are phylogenetically preserved. It is relevant to note, however, that our behavior tests were performed in animals in isolation, and not in a social context. It will be important to study whether these neurons also participate in social behaviors. Additionally, it remains to be tested whether the role of Agrp neurons in feeding and/or repetitive/stereotypic behaviors are influenced by the social context. At present, these studies are extremely challenging to perform in mice (Anderson and Perona, 2014). With the advent of technology and emerging tools to analyze animal behavior, future studies dissecting the role of Agrp neurons (as well as



**Figure 7. NPY<sub>5</sub> Receptor Signaling Is Necessary for AgRP-Neuron-Mediated Behaviors**

(A) Protocol to record home-cage behaviors using CGP71683 (30 mg/kg, i.p.).

(B) Time spent in eating-related behaviors.

(C) Time spent walking.

(D) Total traveled distance.

(E) Time spent digging.

(F) Raster plots showing grooming behavior in individual mouse.

(G) Time spent grooming.

(H) Grooming bouts.

Error bars represent mean  $\pm$  SEM. p values were calculated using two-way ANOVA followed by Holm-Sidak's multiple comparisons test and are reported in the panels. See also Figure S5.

other brain circuits) on behaviors in social settings are of utmost relevance and priority for our understanding of brain function.

Our data also suggest that these ancient brain regions play a role in psychiatric conditions. Specifically, misalignments between environmental cues (peripheral tissue function) and hypothalamic circuits may lead to maladaptive behaviors, including those associated with psychiatric and neurological disorders. Regarding the latter, we suggest that our results have implications for the etiology of anorexia nervosa. Patients suffering from this condition avoid ingesting calories despite the fact that they have elevated activity and a higher physiological state of hunger.

Because hunger signals activate AgRP neurons (Hahn et al., 1998; Liu et al., 2012; Takahashi and Cone, 2005; Yang et al., 2011), we postulate that, in individuals with a vulnerability to develop anorexia nervosa, AgRP neurons may respond to negative energy balance cues in an exacerbated manner and lead to repetitive and compulsive behaviors (Halmi et al., 2003; Matsunaga et al., 1999; Thiel et al., 1995). Future studies are needed to interrogate whether inert differences in AgRP neuronal excit-

ability exist between vulnerable and invulnerable individuals. From this perspective, it is of interest to note that patients with anorexia nervosa have elevated circulating blood levels of AgRP compared to controls (Merle et al., 2011; Moriya et al., 2006) and that AgRP levels are associated with cognitive rigidity in these patients (Sarrar et al., 2011). Because NPY<sub>5</sub> receptor antagonists have been tested in humans (Erondur et al., 2006) and we found it to reverse many AgRP activation-triggered stereotypic behaviors, we suggest that human clinical trials with safe compounds can be initiated for addressing the behavioral aspects of anorexia nervosa as well as other neuropsychiatric diseases with both homeostatic and behavioral components.

## EXPERIMENTAL PROCEDURES

### Mice

All mice used in the experiments were 2–6 months old from both genders. We did not observe differences in the responses of males and females to capsaicin. AgRP-Trpv1 mice were: *AgRP-Cre<sup>Trp1/+</sup>::Trpv1<sup>-/-</sup>::R26-LSL-Trpv1<sup>Gt/+</sup>*; control animals were either AgRP-Trpv1 mice injected with vehicle (3.3% Tween 80 in saline) or *Trpv1<sup>-/-</sup>::R26-LSL-Trpv1<sup>Gt/+</sup>* mice injected with capsaicin. All



animals were littermates (Agrp neuron activated and controls) in the experiments. We did not observe any differences between the two control groups, and therefore, throughout the manuscript we referred to them as “controls.” The following mouse lines were used in this study: *Agrptm1(cre)Low/J*, *Gt(ROSA)26Sortm1(Trpv1,ECFP)Mde/J*, *Trpv1tm1Jul/J*, *Rpl22tm1.1Psam/J*, *Tg(Npy-MAPT/Sapphire)1Rck/J*. All animals were kept in temperature- and humidity-controlled rooms, in a 12/12 hr light/dark cycle, with lights on from 7:00 AM–7:00 PM. Food and water were provided ad libitum unless otherwise stated. All procedures were approved by IACUC (Yale University).

### Immunohistochemistry

Mice were deeply anesthetized and perfused with 0.9% saline containing heparin followed by freshly prepared fixative (paraformaldehyde 4%, picric acid 15%, in PB 0.1M [pH = 7.4]). Brains were post-fixed overnight in fixative. Coronal brain sections (50  $\mu$ m) were washed several times in PB 0.1M (pH = 7.4) and pre-incubated with Triton X-100 for 30 min. Sections were then washed several times and blocked with 2% normal goat serum and incubated with chicken anti-GFP (1:8,000, 4°C, 48 hr; ABCAM), rabbit anti-cfos (1:20,000 at 4°C for 48 hr; Oncogene), and/or mouse anti-HA (1:1,000 dilution at RT for 24 hr; Covance). After, sections were extensively washed and incubated with secondary fluorescent Alexa antibodies (1:500). Sections were mounted, coverslipped, and visualized by a Zeiss microscope or an Olympus Confocal microscope.

### Drugs

Drugs used were: capsaicin (3.33% Tween-80 in PBS; from Sigma), Bicuculline methiodide (in saline; from Sigma), and CGP71683 hydrochloride (in 5% DMSO, 5% Tween-80 in water; from Tocris). All drugs were injected in a volume of 10 ml/kg of body weight intraperitoneally (i.p.).

### Food Intake

For the capsaicin dose-response experiment, mice were acclimated to metabolic chambers (TSE Systems) before recordings. Mice received vehicle or capsaicin (3, 10, and 30 mg/kg, i.p.), and food intake was automatically recorded (see [Movie S1](#)). Alternatively, food intake was manually recorded in single-housed mice. Bedding was changed 24 hr before the experiment, and animals were acclimated for at least 1 week with a minimum quantity of food in the cage to alleviate spillage. On the day of the experiments, food was removed 1 hr before the test and food intake was recorded before and 1 hr after capsaicin injection.

### Electrophysiology

Four-week-old *Agrp-Cre<sup>tm/+</sup>::Trpv1<sup>-/-</sup>::R26-LSL-Trpv1<sup>Gt/+</sup>::NpyGFP<sup>Tg/+</sup>* mice were killed at the beginning of the light cycle, and the arcuate nucleus was sliced into 250  $\mu$ m slices, containing GFP cells. After stabilization in ACSF, slices were transferred to the recording chamber and perfused with ACSF. Basal firing rate was recorded for at least 5 min. The slice was then incubated with a pulse of capsaicin (0.25  $\mu$ M), followed by a washout. Whole-cell current-clamp recording was performed using low-resistance (3–4 M $\Omega$ ) pipettes. The composition of the pipette solution was as follows (in mM): K-glucuronate 125, MgCl<sub>2</sub> 2, HEPES 10, EGTA 1.1, Mg-ATP 4, and Na<sub>2</sub>-phosphocreatine 10, Na<sub>2</sub>-GTP 0.5 (pH 7.3) with KOH. The composition of the bath solution was as follows (in mM): NaCl 124, KCl 3, CaCl<sub>2</sub> 2, MgCl<sub>2</sub> 2, NaH<sub>2</sub>PO<sub>4</sub> 1.23, glucose 2.5, sucrose 7.5, NaHCO<sub>3</sub> 26. After a gigaohm (G $\Omega$ ) seal and whole-cell access were achieved, membrane potential and action potentials were recorded under current clamp at 0 pA. All data were sampled at 3–10 kHz and filtered at 1–3 kHz. Electrophysiological data were analyzed with AxoGraph 4.9.

### Home-Cage Behavior

Four-month-old Agrp-Trpv1 or control female mice were singly housed in their normal home cage 11 days prior to the start of the first behavioral study. Animals were acclimated to handling for 1 week before experiments. The day preceding the behavioral analysis, the mice were given fresh bedding. For a 1 hr acclimation period, cages were placed in front of the cameras of the HomeCageScan system (CleverSys, Reston, VA) and were backlit by IR light panels. Mice were injected with either 10 mg/kg capsaicin or vehicle and recorded

for 1 hr. Food was removed for the acclimation period as well as the analysis period for groups reported as “no food.” Mice in the fasted study were fasted for 16 hr prior to the experiment, and the re-fed group was given food at the time of injection. The NPY<sub>5</sub> receptor blocker (CGP71683 hydrochloride, 30 mg/kg, i.p.) was given to the animals 30 min prior to capsaicin injection. Videos were analyzed with the HomeCageScan software (v3.00).

### Marble-Burying Test

Marble-burying test was as described ([Deacon, 2006](#)) with modifications. Mice were tested (baseline) and randomized to groups. Capsaicin (10 mg/kg, i.p.) was injected immediately before test. Drugs were injected 20 (for bicuculline) or 30 min (for CGP71683 hydrochloride) before capsaicin. Modified marble-burying test was performed in a rat cage containing 40 evenly distributed marbles. Place preference was performed in the same rat cage divided using a separator with an open door. Marble side contained 20 marbles. All studies were performed in cages containing 5 cm of corn-based animal bedding.

### Calorie Restriction

Female mice (9 weeks old) were housed two-by-two to avoid chronic stress due to social isolation. We have used the balanced NIH-41 diet (3.34 kcal/g, protein 16.9%, fat 12.5%, fiber 3.8%, nitrogen-free extract 53.6%, vitamins, minerals) to avoid malnourishment during calorie restriction due to insufficient nutrient levels. Mice received 20% less calories than their ad libitum food intake baseline measurements. The marble-burying test was performed on the last days of the study (a baseline was recorded without injection, and on the next day mice were tested after capsaicin injection). We used the modified marble-burying test with a rat cage containing 40 marbles (as described above).

### DREADD Experiment

Recombinant rAAV5-Ef1a-DIO-hm3D(Gq)-mcherry virus (500 nl from UNC Viral Core) was injected bilaterally into the arcuate nucleus of *Agrp-Cre* male mice (AP = 1.40 mm; DV = −5.90 mm; L =  $\pm$  0.30 mm). Animals were allowed to recover for 3 weeks. All mice were singly housed in their normal home cage 3 weeks prior to the start of the first home-cage behavioral study. Two days preceding the behavioral analysis, the mice were given fresh bedding. Home-cage behaviors were analyzed as above. Mice were injected (i.p.) with either 0.3 mg/kg CNO (n = 7) or saline (n = 4) and recorded during 2 hr with no food available. Mice were later tested for feeding response and showed robust induction of food intake after CNO injection (data not shown). Infection was confirmed by visualizing mCherry in the arcuate nucleus. Clozapine N-oxide (CNO) was from Enzo Life Science.

### Locomotor Activity

Mice were allowed to explore a novel environment (a rat cage, 45  $\times$  24  $\times$  20 cm) for 120 min after capsaicin injection. To test the side effects of the receptor blockers in locomotor activity, animals received an injection of bicuculline methiodide (10 mg/kg, i.p.) or vehicle (PBS) 20 min before experiment. CGP71683 hydrochloride (30 mg/kg, i.p.) or vehicle (5% DMSO, 5% tween-80 in water) were injected 30 min prior to the experiment. Male mice were used in these experiments (n = 25, 3–4 months old) and were allowed to explore the apparatus for 30 min. The experiment was performed under dim light during the light cycle.

### Two-Stages Open-Field Test

The apparatus consists of a Plexiglas open-field (37  $\times$  37  $\times$  37 cm). Mice were first put in the open field for 5 min (“exploratory stage”). Immediately after, mice were returned to their home cages for 2 min. A new object (a cylinder of 5 cm radius and 10 cm high) was placed in the center of the arena. Mice were then returned to the open field for an additional 5 min (“novelty stage”). The room was illuminated with infrared lights and dim red light.

### Elevated Plus Maze and Zero Maze

The plus maze consisted of four elevated arms (40 cm from the floor, 25 cm long, and 5.2 cm wide) arranged at right angles. Two opposite arms were enclosed by 15-cm high walls, and the other two were open (no walls). Male control (n = 8) and Agrp-Trpv1 (n = 11) mice (3–4 months old) were placed on the 5  $\times$  5 cm center section and allowed to explore the apparatus. The zero maze

consisted of an elevated circular platform with two opposite quadrants enclosed and two open, allowing uninterrupted exploration. The apparatus has a 50 cm diameter, 5 cm lane width, 15 cm wall height, and 40 cm elevation (from Stoelting, #68016). Capsaicin (10 mg/kg, i.p.) was injected immediately before the experiments. Experiments were performed during the night cycle of the animals using infrared illumination and dim red light. Mice were recorded for 10 min and tracked using Any-Maze (Stoelting).

### Statistical Analysis

Matlab R2009a, PASW Statistics 18.0, and Prism 6.0 were used to analyze data and plot figures. When homogeneity was assumed, a parametric analysis of variance test was used. The student's *t* test was used to compare two groups. One-, two-way, or two-way with repeated measures ANOVA were used as the other tests unless stated otherwise. When significant, a multiple comparisons post hoc test was used (Holm-Sidak's test). When homogeneity was not assumed, the Kruskal-Wallis nonparametric ANOVA was selected for multiple statistical comparisons. The Mann-Whitney *U* test was used to determine significance between groups. Statistical data are provided in the figures.  $p < 0.05$  was considered statistically significant.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and five movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.024>.

### AUTHOR CONTRIBUTIONS

M.O.D., M.R.Z., and J.B. performed the experiments. All authors designed, analyzed, and interpreted data. M.O.D. and T.L.H. wrote the manuscript.

### ACKNOWLEDGMENTS

We thank Zhong-Wu Liu for electrophysiological recordings. We thank Marya Shanabrough for assistance. T.L.H. was supported by NIH (DP1 DK006850, R01AG040236, and P01NS062686), the American Diabetes Association, the Helmholtz Society (ICEMED), and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq: 401476/2012-0, Brazil). M.O.D. received support from Brain and Behavior Research Foundation, NCATS (UL1 TR000142), and CNPq (487096/2013-4, Brazil). M.R.Z. was partially supported by a Science Without Borders fellowship from CNPq/Brazil.

Received: December 1, 2014

Revised: January 16, 2015

Accepted: January 30, 2015

Published: March 5, 2015

### REFERENCES

Adamah-Biassi, E.B., Stepien, I., Hudson, R.L., and Dubocovich, M.L. (2013). Automated video analysis system reveals distinct diurnal behaviors in C57BL/6 and C3H/HeN mice. *Behav. Brain Res.* 243, 306–312.

Ahmari, S.E., Spellman, T., Douglass, N.L., Kheirbek, M.A., Simpson, H.B., Deisseroth, K., Gordon, J.A., and Hen, R. (2013). Repeated cortico-striatal stimulation generates persistent OCD-like behavior. *Science* 340, 1234–1239.

Amodeo, D.A., Jones, J.H., Sweeney, J.A., and Ragozzino, M.E. (2012). Differences in BTBR T+ tf/J and C57BL/6J mice on probabilistic reversal learning and stereotyped behaviors. *Behav. Brain Res.* 227, 64–72.

Anderson, D.J., and Perona, P. (2014). Toward a science of computational ethology. *Neuron* 84, 18–31.

Aponte, Y., Atasoy, D., and Sternson, S.M. (2011). AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nat. Neurosci.* 14, 351–355.

Arenkiel, B.R., Klein, M.E., Davison, I.G., Katz, L.C., and Ehlers, M.D. (2008). Genetic control of neuronal activity in mice conditionally expressing TRPV1. *Nat. Methods* 5, 299–302.

Atasoy, D., Betley, J.N., Su, H.H., and Sternson, S.M. (2012). Deconstruction of a neural circuit for hunger. *Nature* 488, 172–177.

Barnett, S.A. (1956). Behaviour Components in the Feeding of Wild and Laboratory Rats. *Behaviour* 9, 24–43.

Betley, J.N., Cao, Z.F., Ritola, K.D., and Sternson, S.M. (2013). Parallel, redundant circuit organization for homeostatic control of feeding behavior. *Cell* 155, 1337–1350.

Broberger, C., Johansen, J., Johansson, C., Schalling, M., and Hökfelt, T. (1998). The neuropeptide Y/agouti gene-related protein (AGRP) brain circuitry in normal, anorectic, and monosodium glutamate-treated mice. *Proc. Natl. Acad. Sci. USA* 95, 15043–15048.

Burguière, E., Monteiro, P., Feng, G., and Graybiel, A.M. (2013). Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. *Science* 340, 1243–1246.

Caterina, M.J., Schumacher, M.A., Tominaga, M., Rosen, T.A., Levine, J.D., and Julius, D. (1997). The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* 389, 816–824.

Chaudhri, O., Small, C., and Bloom, S. (2006). Gastrointestinal hormones regulating appetite. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1187–1209.

Clark, J.T., Kalra, P.S., Crowley, W.R., and Kalra, S.P. (1984). Neuropeptide Y and human pancreatic polypeptide stimulate feeding behavior in rats. *Endocrinology* 115, 427–429.

Coll, A.P., Farooqi, I.S., and O'Rahilly, S. (2007). The hormonal control of food intake. *Cell* 129, 251–262.

Deacon, R.M. (2006). Digging and marble burying in mice: simple methods for in vivo identification of biological impacts. *Nat. Protoc.* 1, 122–124.

Dietrich, M.O., and Horvath, T.L. (2009). Feeding signals and brain circuitry. *Eur. J. Neurosci.* 30, 1688–1696.

Dietrich, M.O., Bober, J., Ferreira, J.G., Tellez, L.A., Mineur, Y.S., Souza, D.O., Gao, X.B., Picciotto, M.R., Araújo, I., Liu, Z.W., and Horvath, T.L. (2012). AgRP neurons regulate development of dopamine neuronal plasticity and nonfood-associated behaviors. *Nat. Neurosci.* 15, 1108–1110.

Dube, M.G., Horvath, T.L., Kalra, P.S., and Kalra, S.P. (2000). Evidence of NPY Y5 receptor involvement in food intake elicited by orexin A in sated rats. *Pepptides* 21, 1557–1560.

Erondu, N., Gantz, I., Musser, B., Suryawanshi, S., Mallick, M., Addy, C., Cote, J., Bray, G., Fujioka, K., Bays, H., et al. (2006). Neuropeptide Y5 receptor antagonism does not induce clinically meaningful weight loss in overweight and obese adults. *Cell Metab.* 4, 275–282.

Gerald, C., Walker, M.W., Criscione, L., Gustafson, E.L., Batzl-Hartmann, C., Smith, K.E., Vaysse, P., Durkin, M.M., Laz, T.M., Linemeyer, D.L., et al. (1996). A receptor subtype involved in neuropeptide-Y-induced food intake. *Nature* 382, 168–171.

Gropp, E., Shanabrough, M., Borok, E., Xu, A.W., Janoschek, R., Buch, T., Plum, L., Balthasar, N., Hampel, B., Waisman, A., et al. (2005). Agouti-related peptide-expressing neurons are mandatory for feeding. *Nat. Neurosci.* 8, 1289–1291.

Grove, K.L., Brogan, R.S., and Smith, M.S. (2001). Novel expression of neuropeptide Y (NPY) mRNA in hypothalamic regions during development: region-specific effects of maternal deprivation on NPY and Agouti-related protein mRNA. *Endocrinology* 142, 4771–4776.

Güler, A.D., Rainwater, A., Parker, J.G., Jones, G.L., Argilli, E., Arenkiel, B.R., Ehlers, M.D., Bonci, A., Zweifel, L.S., and Palmiter, R.D. (2012). Transient activation of specific neurons in mice by selective expression of the capsaicin receptor. *Nat. Commun.* 3, 746.

Gyertyán, I. (1995). Analysis of the marble burying response: marbles serve to measure digging rather than evoke burying. *Behav. Pharmacol.* 6, 24–31.

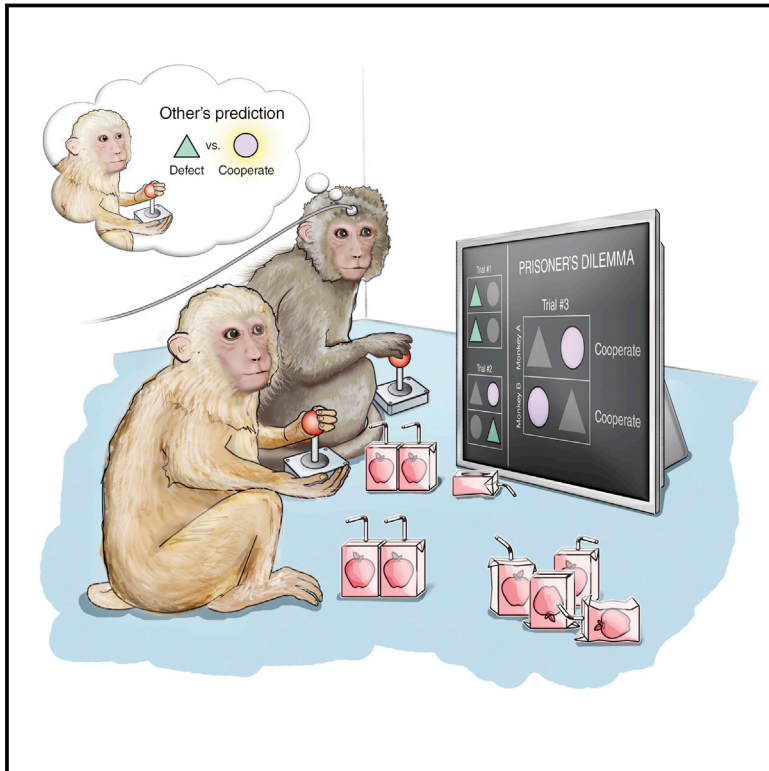
Hahn, T.M., Breininger, J.F., Baskin, D.G., and Schwartz, M.W. (1998). Coexpression of Agrp and NPY in fasting-activated hypothalamic neurons. *Nat. Neurosci.* 1, 271–272.

Halmi, K.A., Sunday, S.R., Klump, K.L., Strober, M., Leckman, J.F., Fichter, M., Kaplan, A., Woodside, B., Treasure, J., Berrettini, W.H., et al. (2003).

- Obsessions and compulsions in anorexia nervosa subtypes. *Int. J. Eat. Disord.* 33, 308–319.
- Horvath, T.L., Naftolin, F., Kalra, S.P., and Leranth, C. (1992). Neuropeptide-Y innervation of beta-endorphin-containing cells in the rat mediobasal hypothalamus: a light and electron microscopic double immunostaining analysis. *Endocrinology* 131, 2461–2467.
- Horvath, T.L., Bechmann, I., Naftolin, F., Kalra, S.P., and Leranth, C. (1997). Heterogeneity in the neuropeptide Y-containing neurons of the rat arcuate nucleus: GABAergic and non-GABAergic subpopulations. *Brain Res.* 756, 283–286.
- Horvath, T.L., Diano, S., and van den Pol, A.N. (1999). Synaptic interaction between hypocretin (orexin) and neuropeptide Y cells in the rodent and primate hypothalamus: a novel circuit implicated in metabolic and endocrine regulations. *J. Neurosci.* 19, 1072–1087.
- Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., and Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* 1, 68.
- Kanatani, A., Mashiko, S., Murai, N., Sugimoto, N., Ito, J., Fukuroda, T., Fukami, T., Morin, N., MacNeil, D.J., Van der Ploeg, L.H., et al. (2000). Role of the Y1 receptor in the regulation of neuropeptide Y-mediated feeding: comparison of wild-type, Y1 receptor-deficient, and Y5 receptor-deficient mice. *Endocrinology* 141, 1011–1016.
- Karvat, G., and Kimchi, T. (2012). Systematic autistic-like behavioral phenotyping of 4 mouse strains using a novel wheel-running assay. *Behav. Brain Res.* 233, 405–414.
- Kheifets, A., and Gallistel, C.R. (2012). Mice take calculated risks. *Proc. Natl. Acad. Sci. USA* 109, 8776–8779.
- Krashes, M.J., DasGupta, S., Vreede, A., White, B., Armstrong, J.D., and Waddell, S. (2009). A neural circuit mechanism integrating motivational state with memory expression in *Drosophila*. *Cell* 139, 416–427.
- Krashes, M.J., Koda, S., Ye, C., Rogan, S.C., Adams, A.C., Cusher, D.S., Maratos-Flier, E., Roth, B.L., and Lowell, B.B. (2011). Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* 121, 1424–1428.
- Krashes, M.J., Shah, B.P., Koda, S., and Lowell, B.B. (2013). Rapid versus delayed stimulation of feeding by the endogenously released AgRP neuron mediators GABA, NPY, and AgRP. *Cell Metab.* 18, 588–595.
- Kyzar, E.J., Pham, M., Roth, A., Cachat, J., Green, J., Gaikwad, S., and Kalueff, A.V. (2012). Alterations in grooming activity and syntax in heterozygous SERT and BDNF knockout mice: the utility of behavior-recognition tools to characterize mutant mouse phenotypes. *Brain Res. Bull.* 89, 168–176.
- Lam, T.K.T., Schwartz, G.J., and Rossetti, L. (2005). Hypothalamic sensing of fatty acids. *Nat. Neurosci.* 8, 579–584.
- Liu, T., Kong, D., Shah, B.P., Ye, C., Koda, S., Saunders, A., Ding, J.B., Yang, Z., Sabatini, B.L., and Lowell, B.B. (2012). Fasting activation of AgRP neurons requires NMDA receptors and involves spinogenesis and increased excitatory tone. *Neuron* 73, 511–522.
- Londei, T., Valentini, A.M., and Leone, V.G. (1998). Investigative burying by laboratory mice may involve non-functional, compulsive, behaviour. *Behav. Brain Res.* 94, 249–254.
- Luquet, S., Perez, F.A., Hnasko, T.S., and Palmiter, R.D. (2005). NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* 310, 683–685.
- Matsunaga, H., Kiriike, N., Iwasaki, Y., Miyata, A., Yamagami, S., and Kaye, W.H. (1999). Clinical characteristics in patients with anorexia nervosa and obsessive-compulsive disorder. *Psychol. Med.* 29, 407–414.
- Merle, J.V., Haas, V., Burghardt, R., Döhler, N., Schneider, N., Lehmkuhl, U., and Ehrlich, S. (2011). Agouti-related protein in patients with acute and weight-restored anorexia nervosa. *Psychol. Med.* 41, 2183–2192.
- Moriya, J., Takimoto, Y., Yoshiuchi, K., Shimosawa, T., and Akabayashi, A. (2006). Plasma agouti-related protein levels in women with anorexia nervosa. *Psychoneuroendocrinology* 31, 1057–1061.
- Ollmann, M.M., Wilson, B.D., Yang, Y.K., Kerns, J.A., Chen, Y., Gantz, I., and Barsh, G.S. (1997). Antagonism of central melanocortin receptors in vitro and in vivo by agouti-related protein. *Science* 278, 135–138.
- Pedrazzini, T., Seydoux, J., Küstner, P., Aubert, J.F., Grouzmann, E., Beermann, F., and Brunner, H.R. (1998). Cardiovascular response, feeding behavior and locomotor activity in mice lacking the NPY Y1 receptor. *Nat. Med.* 4, 722–726.
- Rogan, S.C., and Roth, B.L. (2011). Remote control of neuronal signaling. *Pharmacol. Rev.* 63, 291–315.
- Rossi, M., Kim, M.S., Morgan, D.G., Small, C.J., Edwards, C.M., Sunter, D., Abusnana, S., Goldstone, A.P., Russell, S.H., Stanley, S.A., et al. (1998). A C-terminal fragment of Agouti-related protein increases feeding and antagonizes the effect of alpha-melanocyte stimulating hormone in vivo. *Endocrinology* 139, 4428–4431.
- Sarrar, L., Ehrlich, S., Merle, J.V., Pfeiffer, E., Lehmkuhl, U., and Schneider, N. (2011). Cognitive flexibility and Agouti-related protein in adolescent patients with anorexia nervosa. *Psychoneuroendocrinology* 36, 1396–1406.
- Stanley, B.G., Kyrkouli, S.E., Lampert, S., and Leibowitz, S.F. (1986). Neuropeptide Y chronically injected into the hypothalamus: a powerful neurochemical inducer of hyperphagia and obesity. *Peptides* 7, 1189–1192.
- Takahashi, K.A., and Cone, R.D. (2005). Fasting induces a large, leptin-dependent increase in the intrinsic action potential frequency of orexigenic arcuate nucleus neuropeptide Y/Agouti-related protein neurons. *Endocrinology* 146, 1043–1047.
- Thiel, A., Brooks, A., Ohlmeier, M., Jacoby, G.E., and Schüssler, G. (1995). Obsessive-compulsive disorder among patients with anorexia nervosa and bulimia nervosa. *Am. J. Psychiatry* 152, 72–75.
- Thomas, A., Burant, A., Bui, N., Graham, D., Yuva-Paylor, L.A., and Paylor, R. (2009). Marble burying reflects a repetitive and perseverative behavior more than novelty-induced anxiety. *Psychopharmacology (Berl.)* 204, 361–373.
- Witkin, J.M. (2008). Animal models of obsessive-compulsive disorder. *Curr. Protoc. in Neurosci.*, Unit 9, 30.
- Wolak, M.L., DeJoseph, M.R., Cator, A.D., Mokashi, A.S., Brownfield, M.S., and Urban, J.H. (2003). Comparative distribution of neuropeptide Y Y1 and Y5 receptors in the rat brain by using immunohistochemistry. *J. Comp. Neurol.* 464, 285–311.
- Wu, Q., Clark, M.S., and Palmiter, R.D. (2012). Deciphering a neuronal circuit that mediates appetite. *Nature* 483, 594–597.
- Xu, A.W., Kaelin, C.B., Morton, G.J., Ogimoto, K., Stanhope, K., Graham, J., Baskin, D.G., Havel, P., Schwartz, M.W., and Barsh, G.S. (2005). Effects of hypothalamic neurodegeneration on energy balance. *PLoS Biol.* 3, e415.
- Yang, Y., Atasoy, D., Su, H.H., and Sternson, S.M. (2011). Hunger states switch a flip-flop memory circuit via a synaptic AMPK-dependent positive feedback loop. *Cell* 146, 992–1003.

# Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates

## Graphical Abstract



## Authors

Keren Haroush, Ziv M. Williams

## Correspondence

haroush.keren@mgh.harvard.edu (K.H.),  
zwilliams@mgh.harvard.edu (Z.M.W.)

## In Brief

Cells in the cingulate cortex of primates are able to predict the unknown intentions or state of mind of other individuals and are critically important in enacting cooperative social behavior, a framework that might be relevant for understanding interpersonal, economic, and political decision-making processes in humans.

## Highlights

- Cingulate neurons predict another agent's unknown decisions during social interaction
- Other-predictive neurons are sensitive to social context, but not to expected reward
- Distinct cingulate neurons encode the individual's own decisions to cooperate or defect
- Disrupting cingulate activity selectively inhibits mutually beneficial interactions





# Neuronal Prediction of Opponent's Behavior during Cooperative Social Interchange in Primates

Keren Haroush<sup>1,2,\*</sup> and Ziv M. Williams<sup>1,2,\*</sup>

<sup>1</sup>Harvard-MIT Health Sciences and Technology, Harvard Medical School, Boston, MA 02114, USA

<sup>2</sup>Department of Neurosurgery, MGH-HMS Center for Nervous System Repair, Harvard Medical School, Boston, MA 02114, USA

\*Correspondence: [haroush.keren@mgh.harvard.edu](mailto:haroush.keren@mgh.harvard.edu) (K.H.), [zwilliams@mgh.harvard.edu](mailto:zwilliams@mgh.harvard.edu) (Z.M.W.)

<http://dx.doi.org/10.1016/j.cell.2015.01.045>

## SUMMARY

A cornerstone of successful social interchange is the ability to anticipate each other's intentions or actions. While generating these internal predictions is essential for constructive social behavior, their single neuronal basis and causal underpinnings are unknown. Here, we discover specific neurons in the primate dorsal anterior cingulate that selectively predict an opponent's yet unknown decision to invest in their common good or defect and distinct neurons that encode the monkey's own current decision based on prior outcomes. Mixed population predictions of the other was remarkably near optimal compared to behavioral decoders. Moreover, disrupting cingulate activity selectively biased mutually beneficial interactions between the monkeys but, surprisingly, had no influence on their decisions when no net-positive outcome was possible. These findings identify a group of other-predictive neurons in the primate anterior cingulate essential for enacting cooperative interactions and may pave a way toward the targeted treatment of social behavioral disorders.

## INTRODUCTION

Social interactions are unique from other behaviors in that they inherently require individuals to anticipate each other's unknown intentions and actions. Accordingly, individuals need to consider not only how their decisions affect their own personal outcomes but also how they may affect the outcomes of other individuals in a group and how these individuals may consequently respond. Such interactions, therefore, are not simply governed by the learned sensorimotor contingencies between action and outcome but are rather based on the ability to predict the unknown intentions or "state of mind" of others.

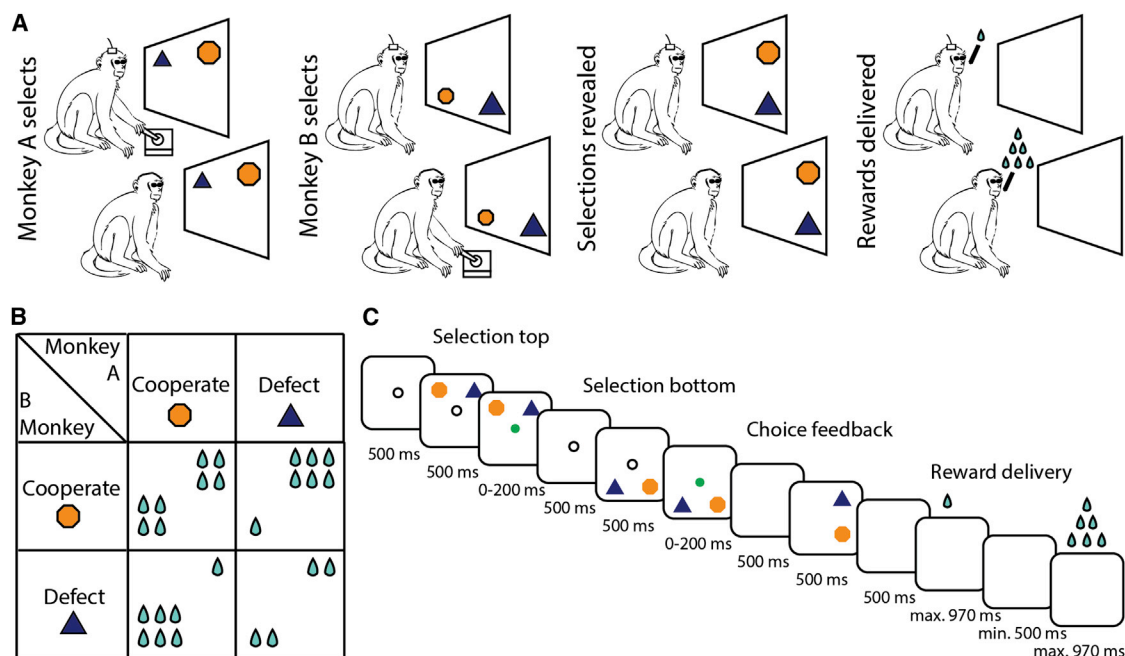
Whether and what neurons encode another's unknown actions and what role these signals play during joint decisions, made independently by two interacting individuals, remain unknown. Prior studies have demonstrated that frontal canonical cells, termed mirror neurons, encode another's known, observable actions, as well as actions performed by the individual himself (di Pellegrino et al., 1992; Rizzolatti and Sinigaglia, 2010). More recently, neurons have been similarly found to encode an-

other's observed receipt of reward (Azzi et al., 2012; Chang et al., 2013; Hosokawa and Watanabe, 2012), as well as monitoring of other's errors (Yoshida et al., 2012, see Discussion). These findings have therefore provided a critical understanding of how another's known and observable actions may be represented at the neuronal level. However, they are distinct from those that may represent another's imminent decisions or intentions, which are fundamentally unobservable and unknown. While cells that predict another's unobservable intended actions have been widely hypothesized, and are a cornerstone of many theories on animal social behavior (Frith and Frith, 1999; Gallese and Goldman, 1998; Rilling et al., 2004; Sanfey et al., 2006; Vogeley et al., 2001), their existence has never been demonstrated.

A second unresolved question is how putative neural signals related to self and other's decisions may affect achieving mutual goals. Mutually beneficial interactions are ubiquitous among social animals (Bshary et al., 2008; Clutton-Brock, 2009; de Waal, 2000; Stephens et al., 2002; Warneken and Tomasello, 2006) and are cardinal to our understanding of socially-guided decisions. While competitive interactions, which allow an individual to profit at the expense of the other, have been previously investigated (Donahue et al., 2013; Hosokawa and Watanabe, 2012; Lee et al., 2005; Seo et al., 2014), the single-neuronal basis of mutually beneficial interactions, favorable to both individuals, have not been explored.

Finally, whereas certain areas may harbor signals that encode elements of social decision-making (Abe and Lee, 2011; Apps et al., 2012; Apps and Ramnani, 2014; Azzi et al., 2012; Behrens et al., 2008; Carter et al., 2012; Chang et al., 2013; Delgado et al., 2005; Donahue et al., 2013; Hampton et al., 2008; Lee et al., 2005; Rilling et al., 2002; Rudebeck et al., 2006; Sanfey et al., 2003; Tomlin et al., 2006; Yoshida et al., 2012), it has not yet been determined what causal contribution neurons in these areas may play in modulating mutual decisions.

A formal framework for studying mutually beneficial joint decisions is by the iterated prisoner's-dilemma (iPD) game (Clutton-Brock, 2009; Rilling et al., 2002; Stephens et al., 2002). This task incorporates two crucial properties: one is that the outcome is contingent upon the mutual concurrent decisions of both individuals, and therefore no one decision guarantees an individual's outcome, and the other is that both decisions can be either concordant or discordant (Camerer, 2003). Therefore, the key to succeeding in the game relies on one's ability to anticipate the other's concurrent, yet unknown intentions. Moreover, this dissociation of self and other decisions, concordant and discordant interactions, and the dissociation between one's decision



**Figure 1. Task Design**

(A) Experimental set-up. The monkeys sat side-by-side, facing a screen. On each trial, they covertly chose, in succession, to cooperate (orange hexagon) or defect (blue triangle). Following delay, both choices were revealed on screen and reward was delivered.

(B) Payoff matrix. Reward outcome for all possible choice combinations. Cooperation and defection were defined operationally by whether mutual benefit or loss is incurred.

(C) Trial timeline. The order in which the monkeys made their selections was randomized on each trial.

and reward, allows one to identify neuronal signals within the population that specifically encode another's yet unknown decisions and importantly dissociate them from those that reflect one's own planned decision and expected reward.

Here, we used a joint-decision paradigm to study mutual decisions in primates and provide evidence of neurons that predict another agent's intentions and modes of cooperation. We specifically focused on the dorsal region of the anterior cingulate cortex (dACC) because of its broad connectivity with frontal and temporal-parietal areas known to be involved in interactive behavior (Behrens et al., 2009; Paus, 2001) as well as its role in encoding social interest in other individuals based on functional imaging (Behrens et al., 2008) and ablative studies (Rudebeck et al., 2006). We find that many dACC neurons encoded the monkey's own decision to cooperate. Furthermore, a substantial and largely distinct group of neurons encoded the opponent monkey's decisions when they were yet unknown. These other-predictive neurons were uniquely sensitive to social context compared to other population cells and encoded no information about the monkey's own decisions or expected reward. At the population-level, dACC neurons reliably predicted the other's decisions with accuracy that remarkably approached those of behavioral decoders when based on prior selections. Finally, transient disruption of dACC activity directly and specifically inhibited mutually beneficial interactions based on prior decisions, but did not affect other decisions based on receipt of reward.

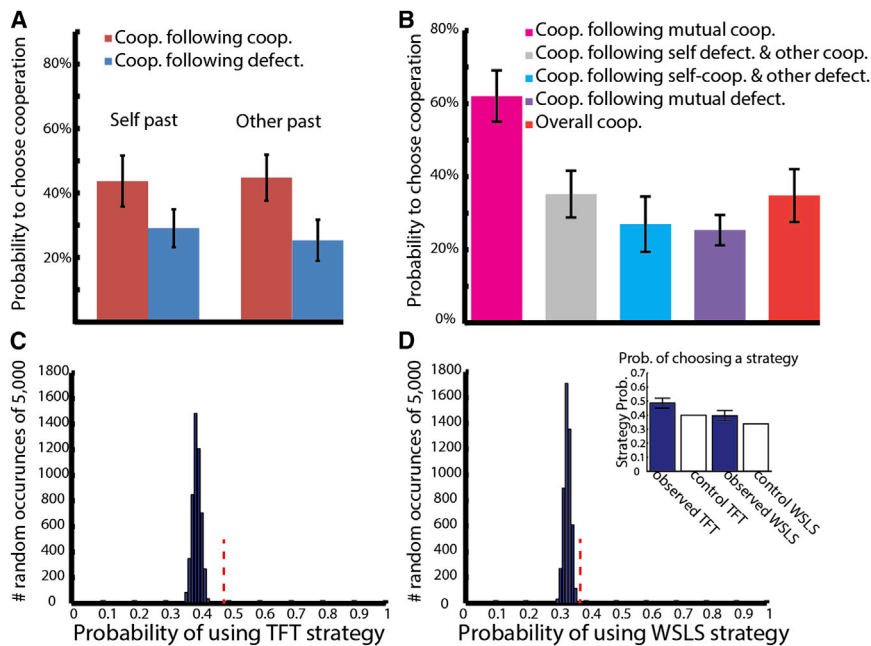
These findings together provide direct examination of how individual neurons represent another's unknown intentions or

covert "state of mind," demonstrate the distinct encoding of other decisions from self-decisions and reward, ascertain the distinct roles that self- and other-encoding cells play in enacting joint decisions between simultaneously interacting animals, and demonstrate a causal link between cingulate activity and the specific enactment of mutually beneficial decisions.

## RESULTS

### Increased Cooperation following Mutual Cooperation

Four pairs of adult male Rhesus monkeys (*Macaca mulatta*) performed an iPD game whereby each animal chose on each trial between two response options over multiple successive trials (Figure 1A). The choice terms, cooperation and defection, were derived from iPD literature (Camerer, 2003). These were defined operationally by the payoff matrix illustrated in Figure 1B and are not referred to here in an anthropomorphic way. If both animals selected cooperation, both received the highest mutual reward whereas if one of the animals defected, that animal received the highest individual reward. The lynchpin of this game, however, was that if neither monkey cooperated, they would both receive a lower reward than if they both chose to cooperate. Accordingly, each individual decision could result in either high or low reward depending on the other's choice, and reward could not be predicted solely from any individual decision. Moreover, since the monkeys performed multiple trials, the decision of an individual to cooperate or defect on one trial may influence the other's subsequent decisions and, therefore, affect the future



**Figure 2. Mutually Beneficial Interactions Increase Cooperation**

(A) Conditional probability of a monkey cooperating given that it cooperated or defected on the preceding trial (left) and conditional probability of a monkey cooperating given the opponent cooperated or defected on the preceding trial (right). Error bars represent SEM.

(B) Probability of selecting cooperation following both monkey's prior mutual selections. Red bar denotes overall cooperation probability. Mutually beneficial interactions led to an increase in subsequent cooperation (this was not evident when playing a computer opponent or in separate rooms, see text).

(C) Probability of following tit-for-tat (TFT) strategy. Histogram shows probability for 5,000 control Monte Carlo realizations of surrogate behavioral data. Red dashed line indicates experimental data value.

(D) Probability of following win-stay-lose-switch (WSLS) strategy. Red dashed line indicates experimental data value. Inset denotes observed data values of both strategies (blue bars), error bars represent SEM, white bars denote mean of surrogate control values.

See also [Figure S1](#).

potential for mutual benefit. Here, we used this setup to differentiate between potential neuronal signals that encoded self-decisions, other-decisions, and expected reward as both monkeys jointly, simultaneously made their own choices.

The monkeys sat side-by-side, facing a screen that displayed different targets representing the choice to cooperate or defect (note, that facial expression observations or eye contact were not possible here by design). Neither monkey saw the other monkey's selection until after they made their own selection plus an additional blank screen delay. Then both selections were revealed on-screen followed by reward ([Figure 1C](#)). To further rule out implicit signals such as auditory cues that may contribute to predictions of the other's decisions, we randomly alternated the order in which monkeys made their selections (see below).

Behaviorally, we find that the monkeys were more likely to select defection over cooperation. The monkeys performed 1,346 trials over seven sessions; they chose defection in 65.3% of trials and cooperation in 34.7% of trials (chi-square = 123.7,  $df = 1$ ,  $p < 10^{-29}$ ). They selected cooperation simultaneously on 17.1% of trials, significantly higher than chance level (chi-square = 44.07,  $df = 1$ ,  $p < 10^{-11}$ ) and both defected on 37.6% of trials, significantly less than chance level (chi-square = 22.27,  $df = 1$ ,  $p < 10^{-6}$ ). Similar to prior observations in humans ([Kuhlman and Marshello, 1975](#); [Rapoport and Chammah, 1965](#)), the monkeys were less likely to cooperate if the other previously defected ( $26\% \pm 6\%$ ;  $2 \times 2$ -chi-square = 56.89,  $df = 1$ ,  $p < 10^{-13}$ ) ([Figure 2A](#)), indicating their understanding of the task by taking into account the other's past action when selecting their own. Moreover, the monkeys were most likely to cooperate if both monkeys cooperated on the preceding trial ( $62.1\% \pm 7.0\%$ ; chi-square = 76.7,  $df = 1$ ,  $p < 10^{-18}$ ) ([Figure 2B](#)), despite the fact that individual reward is maximized if a monkey defects when his opponent continues cooperating (note these choices

did not reflect a simple tit-for-tat response; see [Supplemental Information](#) and [Figure S1](#)). In other words, the monkeys reciprocated mutual cooperation for continued mutual benefit. Finally, we examined the behavioral strategy followed by the monkeys by analyzing specific choice sequences and found that they were significantly different than chance ([Figures 2C](#) and [2D](#); see [Supplemental Information](#)).

### Behavioral Controls

To determine whether the monkeys' choices were affected by social context, i.e., their interaction with another monkey, we repeated the task in the exact same set-up, only now replacing a monkey with a computer opponent ([Chang et al., 2013](#); [Hosokawa and Watanabe, 2012](#)). The computer's choices were determined by the statistics of monkeys' choices on the previous sessions, described above (see [Supplemental Information](#)). We find that the monkeys were less likely to cooperate overall ( $19.1\% \pm 3.9\%$  versus  $34.7\%$ ; chi-square = 161.73,  $df = 1$ ,  $p < 10^{-36}$ ). Moreover, they were less likely to reciprocate cooperation following mutual cooperation ( $14.5\% \pm 3.0\%$  versus  $62.1\%$ ; chi-square = 73.25,  $df = 1$ ,  $p < 10^{-17}$ ) when playing a computer opponent, therefore leading to less mutually beneficial interactions.

To eliminate the possibility that the reduced cooperation resulted from differences in choice selection between the computer model and the behaving monkey, we performed an additional set of social control experiments. Here, the monkeys were placed in two separate rooms so that they could not see the other player or hear each other's licking sounds. In addition, the monkeys' juicers were placed outside the experiment room to eliminate any cues from juicer clicks. Under these conditions, the monkeys performed the same task as before with each other. The monkeys performed a total of 2,344 trials in five experimental sessions. By and large, we find the behavior of the

monkeys in this control to be similar to the behavior found in the computer opponent control. Namely, the overall probability of the monkeys to cooperate under these conditions significantly dropped to 14.2%, compared with 34.7% when playing together (chi-square = 432.08,  $df = 1$ ,  $p < 10^{-95}$ ). Furthermore, we did not observe the increased cooperation following mutual cooperation that was a signature of the monkeys' behavior when playing each other in the same room. Namely, the probability of cooperating following a mutual cooperation trial dropped to 17.4% compared with 62% when playing in the same room (chi-square = 38.76,  $df = 1$ ,  $p < 10^{-9}$ ). This value closely matches the computer control value of 14.5% (not significant [n.s.] difference). Therefore, the effect of social context on the behavior of the monkeys is corroborated by these two independent control experiments (i.e., computer control and other room control).

As noted above, the monkeys demonstrated their understanding of the task by taking into account past joint decisions when selecting their own. However, to further confirm that the monkeys understood the relationship between their choices and payoff, the monkeys performed an additional control version of the task in which they were presented with the same choices as before, but could now see the other's selection before responding (see [Supplemental Information](#)). We find that, on trials in which the other monkey first defected, the monkey maximized reward by subsequently selecting defection on  $90.7\% \pm 2.2\%$  of trials (i.e., within the same trial when no mutual beneficial outcome was possible). This held true even if the other monkey cooperated on the preceding trial ( $95.0\% \pm 3.0\%$ ). In other words, the monkeys did not reciprocate a prior offer of cooperation if they knew their opponent defected on the present trial. This did not reflect a simple reward maximization behavior (see [Supplemental Information](#)).

### Single Neuronal Encoding of Another Individual's Unknown Decisions

We recorded 363 neurons in the dACC in two of the four monkeys during task performance. Of these, 185 neurons significantly responded to the task (stepwise linear regression of neuronal firing rate with both monkeys' current and past decisions as predictor variables, corrected for comparisons across pre- and post-selection periods) ([Figures 3A–3D](#) and [S2](#); [Table S1](#); [Experimental Procedures](#); [Supplemental Information](#)). In total, 24.3% of neurons encoded the monkey's own choices on the current trial; 15.7% responded differentially to choosing cooperation versus defection during the pre-selection period (immediately before the monkey's selection) while 11.4% responded differentially during the post-selection period (immediately after the monkey's selection;  $p < 0.05$ ) ([Figure 3A](#)). There was a 2.33-fold  $\pm$  0.26-fold change in absolute activity between cooperation and defection when considered across all such neurons ( $p < 0.05$ ). While the sign of the modulation of neural activity was similar in most neurons when the monkeys chose to defect, responses were more variable across neurons when the monkeys chose cooperation. Approximately half of these neurons (54.7%) had an increase of activity whereas the other half presented a decrease in activity ([Figure 3C](#), left panel). In other words, many dACC neurons encoded the monkey's decision to cooperate or defect.

The key for succeeding in this game was the ability to anticipate the other monkey's concurrent decisions. Analyzing neural activity during the time when monkeys were still unaware of the other's concurrent selection, we found that the activity of many neurons was modulated by the other monkey's yet unknown upcoming choice. A total of 32.4% of neurons demonstrated significant differences in activity when the other monkey concurrently selected cooperation versus defection. Most of these (27.6%) encoded the opponent's unknown choice during the post-selection period (but prior to being informed of the other's response) and 7% during pre-selection period ( $p < 0.05$ ) ([Figure 3B](#)). There was a 1.81-fold  $\pm$  0.07-fold change in absolute activity between other's cooperation and defection when considered across all such neurons ( $p < 0.05$ ) ([Figure 3C](#), right panel; note that the total number of neurons encoding current decisions was larger when considering past responses; see [Supplemental Information](#) and further below).

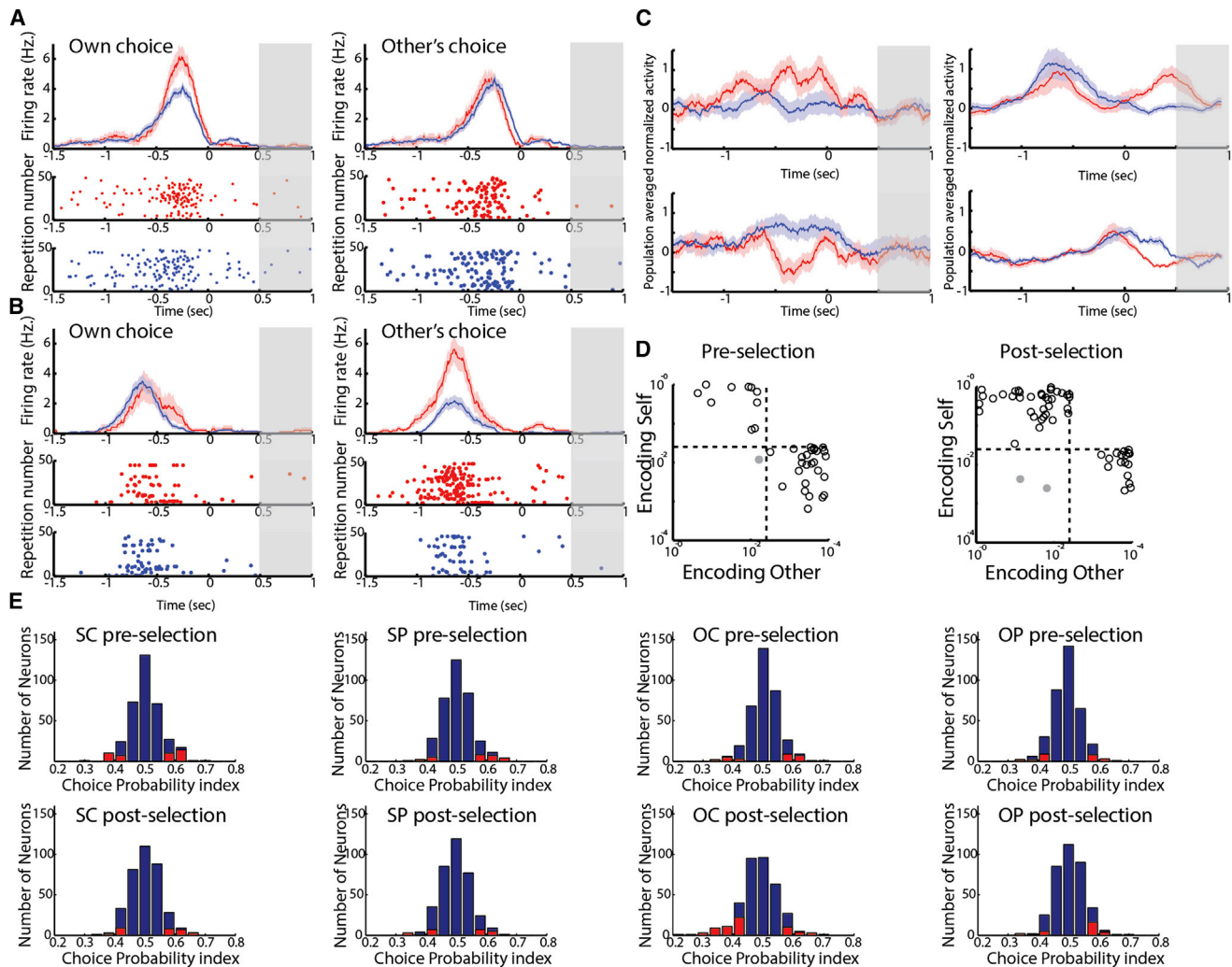
Neurons encoding the opponent monkey's choices and neurons encoding the monkey's own choices demonstrated little overlap with each other ([Figure 3D](#)). Only 4.3% of neurons responded to both the monkey's own decisions as well as the opponent's planned decisions. This was significantly lower than chance level, i.e., that expected by a product of the individual probabilities of encoding self and other (expected: 7.9%, chi-square = 4.97,  $df = 1$ ,  $p < 0.026$ ). This suggests that self and other related computations were carried out by largely distinct neuronal populations ([Figures S3](#) and [S4](#); [Supplemental Information](#)).

To further delineate and confirm the response characteristics of these neurons, we applied three additional approaches to re-analyze the data. First, we performed a choice probability (CP) index analysis examining the trial-by-trial encoding of single neuronal responses. CP index analysis results closely matched the stepwise regression results (35.7% of task responsive neurons had a significant CP index for encoding the other's choice post-selection, and 21.6% had a significant CP index for encoding self-decision pre-selection) ([Figures 3E](#) and [S5A](#); [Supplemental Information](#)). Second, we performed an Akaike Information Criterion (AIC) analysis, which penalizes models containing multiple terms, to complement the term selection process in the stepwise linear regression ([Figures S5B–S5E](#); [Tables S2A](#) and [S2B](#)). Finally, we performed an unsupervised population analysis in the form of a mixture of linear regression models to test in a more unbiased fashion the behavioral factors to which neurons responded at the population level ([Figures S6A–S6F](#)). These analyses confirmed the existence of self and other encoding neurons and the prominence of other-predictive neurons in the dACC and further demonstrate that our findings based on the neuronal data were reproducible across statistical methods (see [Supplemental Information](#)).

### Neurons Predicting the Other's Unknown Decision Are Sensitive to Social Context

To test the direct effect of social context on neural encoding, we recorded a total of 164 additional neurons from the dACC during the social control experiment in which the monkeys played together but in separate rooms. Of these, 84 neurons were found task-responsive using the same stepwise regression analysis as





**Figure 3. Distinct dACC Neurons Encode Self and Other's Decisions**

Peristimulus histograms as mean firing activity  $\pm$  SEM and raster plots for individual neurons. Cooperation trials are denoted in red and defection in blue. Time zero denotes monkey's own selection.

(A) Left: an example of a neuron that encoded the monkey's own current decision to cooperate or defect. Right: the same neuron did not encode the opponent's yet unknown decision. Gray bar indicates the time when both decisions were revealed to the monkeys (on half of trials; see text).

(B) Example of a neuron that encoded the opponent monkey's yet unknown decision to cooperate or defect (right), but did not encode the monkey's own current decision (left).

(C) Population responses based on the monkey's own current decisions for neurons that had a significantly higher activity during self-cooperation versus self-defection (top left) and significantly lower activity during self-cooperation versus self-defection (bottom left); and population responses for neurons that had significantly higher activity during other-cooperation versus other-defection (top right) and significantly lower activity during other-cooperation versus other-defection (bottom right).

(D) Functional partitioning within the population between neurons encoding the monkey's own current decisions and the opponent's yet unknown decisions. Log-log-scale scatter plots of individual neurons p values obtained from the regression analysis during pre- (left) and post-selection (right) periods (only significant neurons are shown). Dashed lines denote significance thresholds. Gray points denote neurons that significantly encoded both the monkey's own decisions and the opponent's decisions.

(E) Neurons with significant modulation based on choice probability (CP) analysis. Top row: pre-decision time period, bottom row: post-decision time period. Columns from left to right correspond to different behavioral variables (SC, self-current; SP, self-past; OC, other-current; OP, other-past). Red bars indicate significant neurons as obtained by bootstrap estimate.

See also [Figures S2, S3, S4, S5, and S6](#) and [Tables S1, S2A, S2B, and S3](#).

above ( $p < 0.025$  for any main or interaction effect, either during the pre or post selection period; see [Table S3](#)). We found that only 14.3% of task responsive cells predicted the other's choice,

significantly less than the 27.6% observed in the main task (chi-square = 7.42,  $df = 1$ ,  $p < 0.006$ ; post-decision). In contrast, a significantly larger fraction of task-responsive neurons encoded

the monkey's own decision in the separate room control (21.4% during the pre-selection period and 26.2% during the post-selection period, compared to 15.7% and 11.4% respectively in the main task; pre-selection: chi-square = 2.083,  $df = 1$ ,  $p = 0.149$ ; post-selection: chi-square = 18.193,  $df = 1$ ,  $p < 0.00002$ ). One possible explanation for the higher number of neurons encoding the monkey's own decisions is that there were more trials recorded per session during the separate room control. However, if this was the only factor, we would also expect to have a concurrent increase in the number of other-predictive neurons, which was not the case. Moreover, the increase in neurons encoding self-decisions indicates that the drop in other-predictive neurons was not simply due to a difference in the raw number of overall cooperation/defection trials. Therefore, this considerable reduction in the fraction of other-predictive neurons indicates that other-predictive neurons are significantly and selectively sensitive to social context.

### Neurons Encoding the Other's Unknown Decisions Do Not Encode Expected Reward

While certain cingulate cells are known to encode received and expected reward (Seo and Lee, 2007; Sheth et al., 2012; Williams and Eskandar, 2006), cells encoding self or other decisions were largely distinct from those that encoded expected reward. An important feature of the iPD game is that it enables one to dissociate neuronal signals encoding self and other decisions from those related to expected reward. Specifically, the monkey's own choice alone cannot guarantee a high or low reward. Therefore, predicting one's own reward inherently requires an accurate prediction of the opponent's yet unknown selection. Nonetheless, to demonstrate more directly that the activity of cells predicting other-decisions is not explained by encoding of expected reward, we provide four lines of evidence based on examining the neuronal responses across multiple behavioral outcome contingencies.

First, we directly examined the encoding of expected reward during the decision period. We found that none of the other-predictive neurons was significantly modulated by self-reward across all four reward contingencies determined by the payoff matrix (see Supplemental Information for statistical tests). Second, we examined the differences in firing rate modulation between encoding of other decision and encoding of self-reward across the recorded population. We found that the firing rate modulation of other-predictive neurons was strong and significantly different from the general population when considering differences in the other's choice to cooperate or defect (Figure 4A), but not when aligning trials according to differences in the monkey's own expected reward, i.e., comparing trials in which the monkey cooperated or defected when the other choose to defect (Figure 4B) and when the other chose to cooperate (Figure 4C). Note that while we did find neurons in the dACC that showed strong modulation to self and other reward (as previously reported by Azzi et al., 2012; Chang et al., 2013; Hosokawa and Watanabe, 2012), these were distinct from the other-predictive neurons (Figure S7A; Supplemental Information). Third, we examined the reward feedback period itself, as it may have been possible that other-predictive neurons only encode reward weakly during the decision period when outcome

is uncertain, but are more strongly modulated by reward when it is certain or known. However, we found that this was not the case (Figure 4D). In fact, compared to other cingulate cells, which overall demonstrated an enhanced modulation to expected reward during feedback, other-predictive neurons demonstrated a slight, non-significant reduction in modulation (Figure 4E). Finally, to test whether other-predictive neurons could be simply sensitive to raw difference in amount of reward irrespective of choice, we repeat the comparison between feedback time modulation and decision time modulation, but for the contingency that yielded the maximal difference in reward, and find no difference in modulation of the other-predictive neurons (Figure 4F).

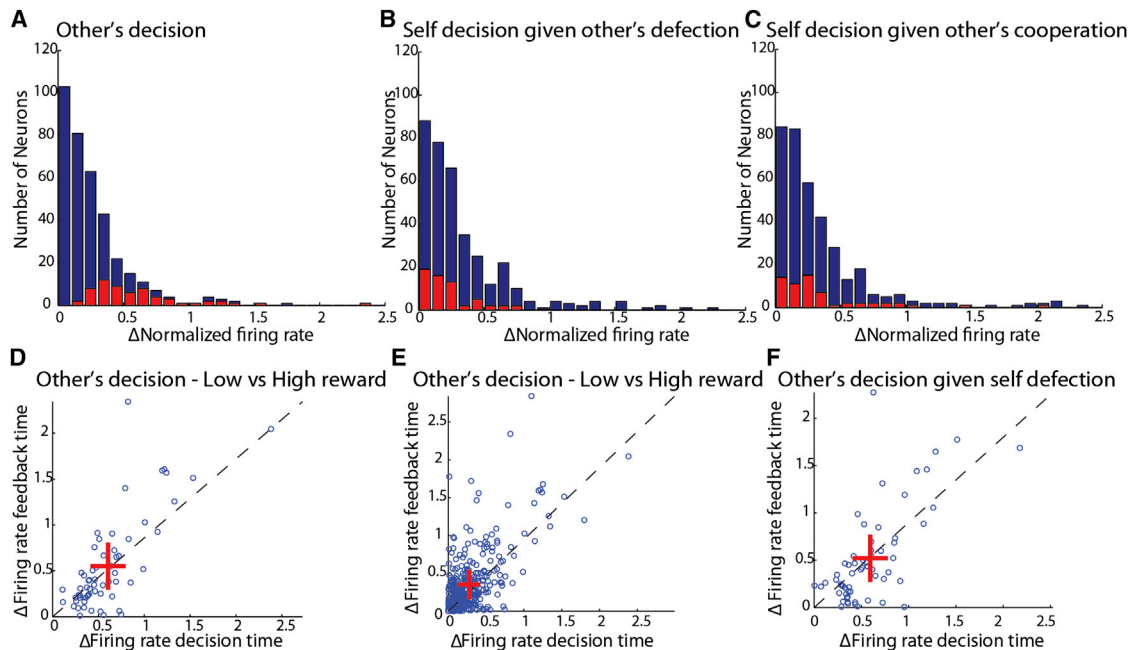
In summary, we demonstrate that the response properties of other-predictive neurons were not explained by simple encoding of the monkey's own expected reward (see Supplemental Information). These results are further bolstered by the finding above that other-predictive neurons encoded no significant information about self-decisions and that they were highly sensitive to social context compared to other population cells.

### dACC Populations Accurately Predict the Other's Decisions on a Trial-by-Trial Basis

Activity in the dACC was significantly predictive of self and other's choices on a trial-by-trial basis when considered across the entire population (Figures 5A and 5B). We constructed a linear decoder to predict the monkeys selections based on population activity (see Supplemental Information). Evaluating model performance on validation trials not used for model training, we find that cingulate populations predicted up to  $66.1\% \pm 0.9\%$  of the recorded monkey's own current choices (multivariate analysis of variance [MANOVA],  $p < 10^{-4}$ ), with predictions being most pronounced in the pre-selection period (Figure 5C). Surprisingly, population activity correctly predicted the other monkey's yet unknown choices on up to  $79.4 \pm 1.1\%$  of trials (MANOVA,  $p < 10^{-5}$ ), with predictions being most pronounced in the post-selection period (Figure 5D). Prediction of other's unknown choices was significantly more accurate than prediction of monkey's own current choices (paired  $t$  test,  $p < 10^{-5}$ ).

To more directly examine the role that the cells selected as other-predictive neurons by the regression analysis play in population decoding of the other's yet unknown decision, we next ran the decoder using only this subset of the neuronal population. We find that the accuracy of predicting the other monkey's decision was not affected and remained up to  $78.1\% \pm 0.8\%$  (MANOVA,  $p < 10^{-9}$ ) correct, despite the fact that the decoder had access to far less cells. However, the accuracy of decoding the monkey's own decisions drastically dropped and was only up to  $54.7\% \pm 0.9\%$  (MANOVA,  $p = 0.37$ , n.s.). These specific effects found in restricting the analysis to this subset of neurons further support the above ascribed role of other-predictive neurons, as well as the functional distinction between these cells and those that encode the monkey's own selections.

Finally, we considered whether implicit cues between the two monkeys could explain these predictions. Note that an important aspect of the task design was that the monkeys made their selections in random temporal order before their responses were revealed. Accordingly, we tested the population predictions



**Figure 4. Other-Predictive Neurons Do Not Encode the Monkey's Own Expected Reward as Shown across Multiple Reward Contingencies**

(A) Histogram of normalized difference in firing rate between trials in which the other monkey defected versus cooperated. Red bars indicate other-predictive neurons. Blue bars indicate the full population. The distributions were statistically different.

(B) Histogram of normalized difference in firing rate between trials in which the monkey chose defection versus cooperation, conditioned on the other choosing defection. Red bars indicate other-predictive neurons. Blue bars indicate the full population. No significant difference was found between distributions.

(C) Histogram of normalized difference in firing rate between trials in which the monkey chose defection versus cooperation, conditioned on the other choosing cooperation. Red bars indicate other-predictive neurons. Blue bars indicate full population. No significant difference was found between distributions.

(D) Scatter plot of firing rate difference between trials in which the other defected versus cooperated, for firing rate during decision time (x axis) and feedback time (y axis) in other-predictive neurons. There is no increase in differential activity when reward is known. Crosses represent mean  $\pm$  SEM.

(E) Scatter plot of firing rate difference between trials in which other defected versus cooperated, for firing rate during decision time (x axis) and feedback time (y axis) in the full population. Here, there was a significant increase in differential activity when reward is known.

(F) Scatter plot of firing rate difference between trials in which the monkey chose defection versus cooperation, conditioned on other's defection, for firing rate during decision time (x axis) and feedback time (y axis) in other-predictive neurons. Here, there is no increase in differential activity when reward is known. See also Figure S7.

when considering only trials in which the monkey played first, i.e., when the other monkey hadn't yet made his selection. We found that predictions of other's unknown choices maintained high accuracy (up to  $70.7\% \pm 0.8\%$ ) and similar accuracies were found when considering only trials in which the monkey played second ( $68.5\% \pm 7.2\%$ ), ruling out the possibility that prediction is an artifact of an implicit signal disclosing the other monkey's choice. Note lower accuracy was expected due to using half the number of trials.

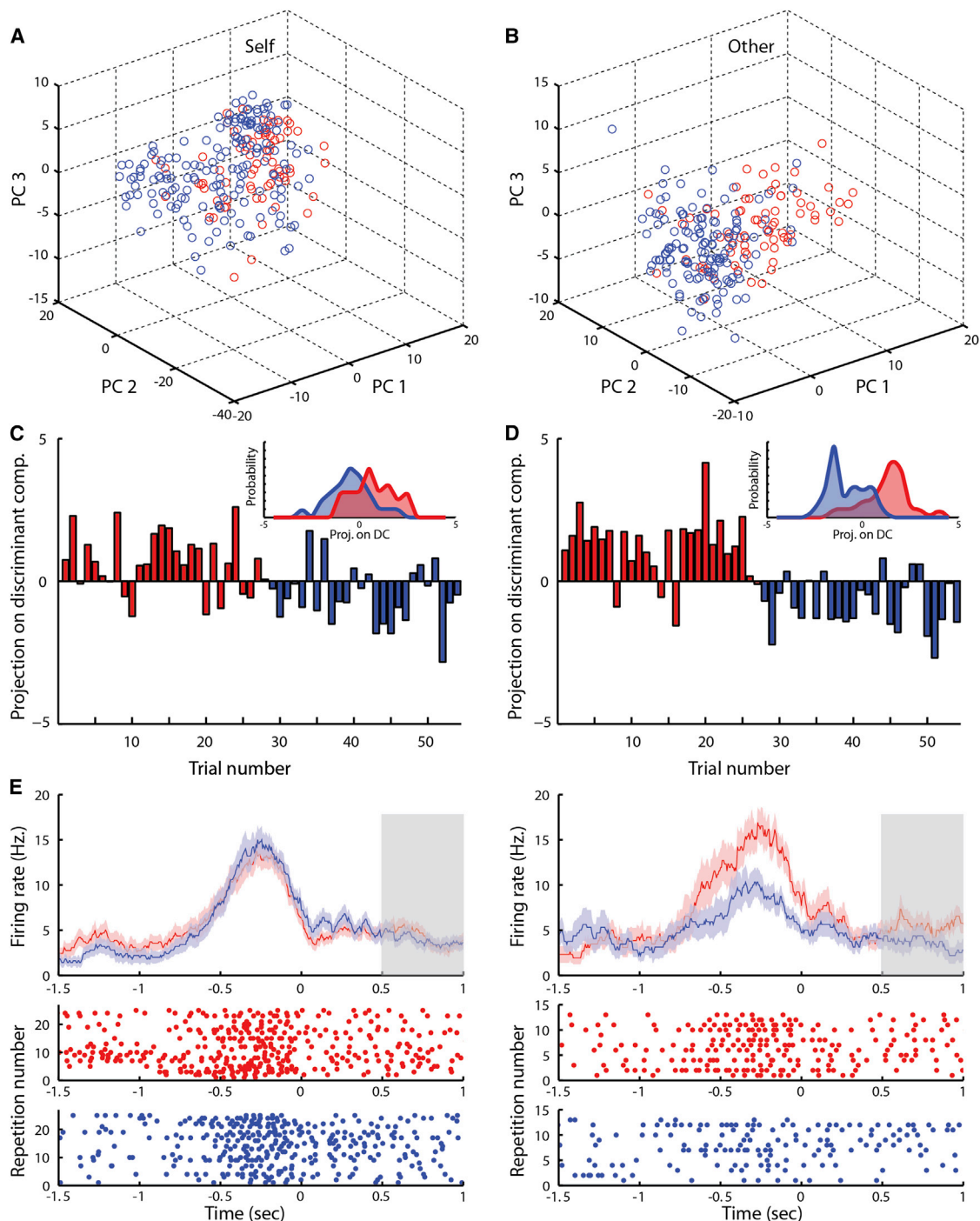
### Behavioral Trial-by-Trial Decoders

To search for a possible basis for neural prediction of the other's concurrent selections, we examined predictions based on both monkeys' prior behavioral history. Using a locally-optimal classification model considering the monkeys' selections four trials back, we estimated on validation trial data the accuracy of predicting the opponent monkey's unknown concurrent choices. We find that model prediction accuracy was up to 79.8%, similar to neuronal decoding (similar accuracies were found for predicting self-selections, see [Supplemental Information](#)). To further explore the behavioral basis of the neuronal predictions of

other's decisions, we tested trial-by-trial correlation between the behavioral and population-activity predictors, revealing significant correlations based on both monkeys' past selections ( $r = 0.31$ ,  $p < 0.0003$ ). These correlations of other's predictions were not evident when behavioral predictions were based on only a single monkey's past decisions or reward (see [Supplemental Information](#)). This suggests that population predictions were based on the prior choices of the two monkeys rather than any individual's past response or reward.

### Neurons Keeping Track of Past Interactions

Consistent with the above findings, we find that many neurons within the population kept a dynamic record of the monkeys' prior selections. [Figure 5E](#) illustrates such a neuron; when the monkey chose to currently defect (left panel), responses did not differ when, on the preceding trial, the opponent chose to defect versus cooperate. In contrast, when the monkey himself cooperated (right panel), neuronal activity was significantly inhibited on trials in which the opponent previously defected (i.e., the monkey cooperated despite the opponent previously defecting) compared to those in which the opponent cooperated



**Figure 5. Trial-by-Trial Population Prediction of the Other's Yet Unknown Decision**

(A and B) Principal component (PC) analysis over a sample session. Plotted in first three PC space, each circle represents the activity of all neurons recorded simultaneously on a single cooperation (red) or defection (blue) trial (see [Supplemental Information](#)).

(A) Self-current pre-decision activity.

(B) Other's-concurrent (yet unknown) post-decision activity.

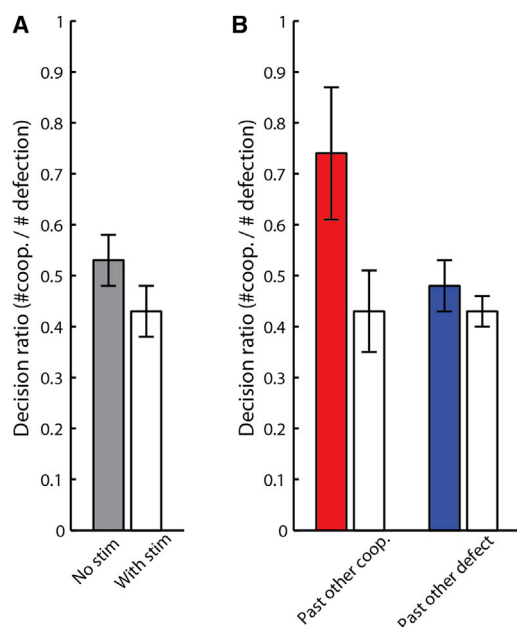
(C and D) Linear decoding model. Each bar represents projection of the activity of all simultaneously recorded neurons during a single trial on first discriminant component (color code above). Positive values predict cooperation and negative defection. Insets (top right) plot distribution of projection values for cooperation (red) and defection (blue).

(C) Self-current pre-selection projection.

(D) Other's-concurrent projection during post-selection.

*(legend continued on next page)*





**Figure 6. dACC Stimulation Selectively Inhibits Mutually Beneficial Interactions**

White bars represent stimulation trials.

(A) Proportion in which the monkeys chose cooperation over defection  $\pm$  SEM (decision-ratio of 1 indicates equal proportion of selecting either).

(B) Decision-ratio given the opponent's past decisions to cooperate (left), or defect (right).

(i.e., reciprocating opponent's preceding cooperation). In addition we found neurons that differentially encoded the joint outcomes on preceding trials (see Figure S7B and Supplemental Information for further details).

### Cingulate Disruption Selectively Inhibits Mutually Beneficial Interactions

Given the above physiological findings, we next investigated whether disruption of the dACC may influence the monkeys' mutual choices. A series of electrical pulses was delivered to the dACC on half of 3,026 randomly selected trials in blocks (1,000 ms triggered at image presentation; 100  $\mu$ A, 200- $\mu$ s biphasic pulse durations with cathodal phase leading; see Supplemental Information).

Stimulation had a significant and selective effect on the monkeys' decisions. Here, we defined the "decision-ratio" as the number of trials in which the monkey selected cooperation over defection (i.e., a ratio of 1 indicates equal selection of cooperation versus defection). When no stimulation was given, the decision-ratio was 0.53 (corresponding to 34.7% cooperation, as also found in the main task). When stimulation was administered, the decision-ratio dropped to 0.43, i.e., monkeys were less likely to cooperate when stimulated ( $t(6) = 3.18$ ,  $p < 0.01$ )

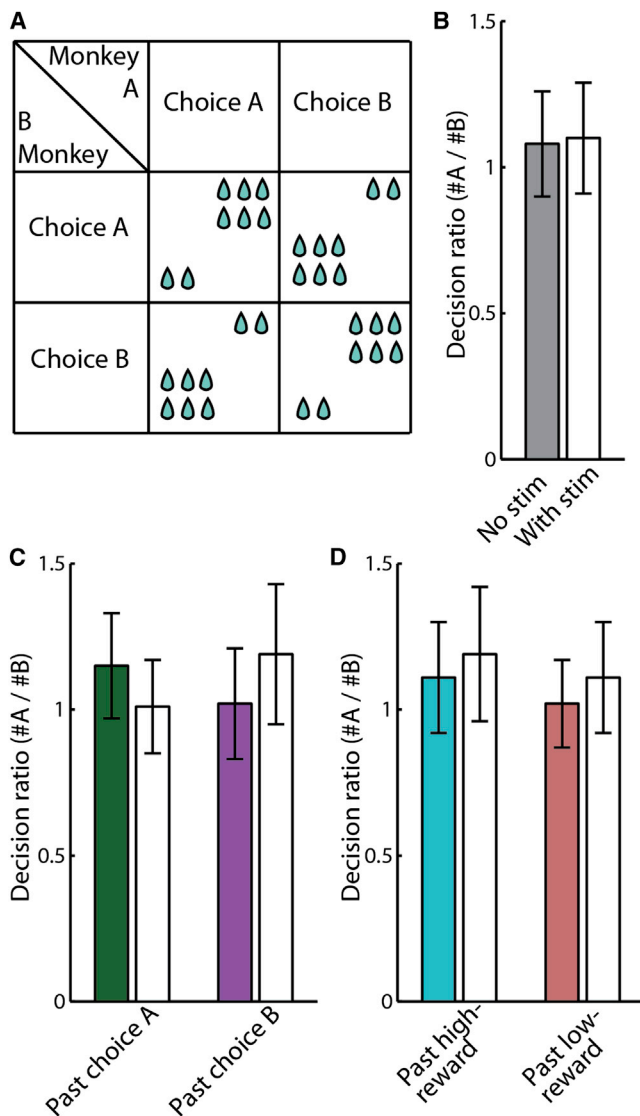
(Figure 6A). This effect was highly dependent on the opponent monkey's preceding selection. When the opponent previously cooperated and no stimulation was given, the decision-ratio was 0.74, meaning that monkeys were more likely to choose cooperation if the opponent previously chose cooperation. However, during stimulation, following opponent's cooperation, the decision-ratio significantly dropped to 0.43 ( $t(6) = -5.57$ ,  $p < 0.0007$ ) (Figure 6B). In contrast, following opponent's defection when no stimulation was given, the decision ratio was 0.48 and, when stimulation was given, it was 0.43 ( $t(6) = -1.12$ ,  $p = 0.15$ ). In other words, stimulation had no effect on the monkey's current decision if the opponent previously defected, but when the opponent previously cooperated, stimulation reduced the decision-ratio to a level equal to the opponent previously choosing defection. Moreover, stimulation had no effect on risk behavior since the rate of cooperation when the other monkey defected on the preceding trial was not affected by stimulation (even though the risk of cooperation under such a condition is higher; i.e., the probability of the opponent to defect following defection is twice higher than following cooperation).

Finally, to further confirm that stimulation did not simply affect decisions based on past reward, we employed a zero-sum game task in which monkey's reward was contingent on the other's response, but individual profit was always at the expense of the other and no mutual positive outcome was possible (i.e., playing under Pareto optimality conditions) (Nash, 1950). We found no effect of stimulation on monkeys' choices during the zero-sum game, based either on the monkeys' preceding selection or preceding receipt of reward (Figure 7; Supplemental Information). Taken together, we conclude that stimulation in the dACC abolished specifically the incorporation of recent positive interactions, rather than any past interaction, into the monkey's own current decision, resulting in less mutually beneficial interactions.

### DISCUSSION

Identifying neurons that reflect another individual's covert intentions or "state of mind" has been a long sought goal in neuroscience and a central proposed tenet of social decision making (Frith and Frith, 1999; Rilling et al., 2004; Sanfey et al., 2006; Vogeley et al., 2001). Here, we discover neurons that selectively encode another individual's yet unknown decisions during joint interactions. We confirmed that no explicit cues were relayed between the two monkeys during the task by using alternating trials in half of which the monkey from which we obtained recordings played first. We also demonstrated reliable population predictions of the other's decisions even on trials in which the other monkey had not yet made his selection. Remarkably, other-predictive cells during joint interactions constituted over a third of the cingulate task-responsive population and were more prevalent than cells encoding the monkey's own present selections. Notably, other predictive neurons were highly sensitive to social

(E) Peristimulus histograms as mean firing activity  $\pm$  SEM (top) and raster plots of a neuron encoding the monkey's own current decision during the pre-selection period and modulated by the other's past decision. Trials separated according to monkey's own current decision to defect (left) or cooperate (right) and opponent's decision on a preceding trial to cooperate (red) or defect (blue; see text). Time zero denotes monkey's own selection. Gray bar indicates feedback period.



**Figure 7. Stimulation Has No Effect when No Mutually Beneficial Interactions Were Possible**

(A) Zero-sum game payoff matrix.

(B–D) Bars represent the decision-ratio on stimulated (white) and non-stimulated (colored) trials during the zero-sum game (see [Supplemental Information](#)). Error bars represent SEM. (B) Overall decision-ratio. (C) Decision-ratio was not affected by opponent's selection of choice A (left) or choice B (right) on the preceding trial. (D) Decision-ratio was not affected by the monkey's own past reward. Left bars: the monkey previously received a high reward. Right bars: the monkey previously received a low reward.

context and were not modulated by self-decisions or expected reward. Consistently, population predictions of the opponent's selections were more accurate than those reflecting the monkey's own selections and, in fact, predicted the other monkey's decisions with accuracies that were near optimal compared to behavioral decoders that considered both monkey's past behaviors. Taken together, these findings provide understanding of the population partitioning by which individual neurons in the pri-

mate cingulate cortex encode information about other social agents.

Game theory provides a framework for dissecting specific aspects of joint decision making, namely the contributions of self and other choices to shared outcome. Signals related to another's yet unobservable actions, in particular, are a distinct feature of mutual interactions in that one participant's concurrent decision affects the other's outcome and therefore inherently requires each participant to anticipate the other's intentions or state of mind.

These predictive signals are fundamentally distinct from previously reported neurons which reflect another animal's known and observable actions. These include canonical mirror neurons that reflect one's observed behavior and do not distinguish between self and other (di Pellegrino et al., 1992; Rizzolatti and Sinigaglia, 2010), neurons that encode another's observed receipt of reward (Azzi et al., 2012; Chang et al., 2013), and neurons that monitor other's observable errors (Yoshida et al., 2012). Importantly, the prediction neurons reported here are distinct from the findings of the latter study, in which neurons monitored the other's errors while the monkeys explicitly observed each other's selections on the same shared task (with each monkey alternating between actor and observer every other trial) (Yoshida et al., 2012). Moreover, encoding of the other's error occurred within the monkeys' movement time window (<200 ms before other's response) and in a setup which allowed them to directly observe each other's movement-preparatory cues. Here, decisions were made jointly, the other's decisions were inherently unobservable and unknown, and their neural encoding could be found many seconds before the other monkey made a selection.

A central feature of non-competitive games such as iPD is that no particular decision guarantees a high or low reward and different outcomes can be experienced either mutually or individually. This dissociation enabled us to examine the computations that contributed to self and other predictions and differentiate them from those that contribute to the encoding of reward outcome. More importantly, it allowed us to examine what particular computations were associated with interactions that were mutually beneficial compared with those that were not. For instance, the monkeys were almost twice as likely to cooperate if they both cooperated on the preceding trial, indicating an intention to reciprocate mutual cooperation. Here, we find that neurons that encoded a monkey's decisions largely did not encode his past or future receipt of reward even though, in combination, these neural signals could be used to predict the monkey's shared outcome. Many neurons, however, were also highly modulated by the two monkey's prior selections. For example, certain neurons differentially encoded the monkey's present decision to cooperate, based on the other monkey's preceding selection of cooperation or defection. Similarly, at the population level, neuronal predictions strongly correlated with predictions made by the behavioral decoder when considering both monkeys' past selections, indicating that neural predictions were based on the past interaction of both individuals.

Consistent with these physiological findings, we observed that disruption of the dACC by stimulation reduced the monkey's

likelihood of cooperation, an effect which was most evident when the opponent cooperated on the preceding trial. Stimulation therefore affected reciprocation of the other's cooperation, but did not affect the animal's ability to incorporate any past decision or outcome since no effect was observed when the opponent defected on the previous trial, or when testing the monkey's decisions in a zero-sum game. This is consistent with previous studies employing a computer opponent in zero-sum games that showed that the dACC does not differentially encode the monkey's decisions during such interactions (Donahue et al., 2013; Seo and Lee, 2007). Therefore, during joint interactions, the dACC specifically mediated mutually beneficial decisions based on the recent history of the interaction.

The monkeys were clearly affected by the social context of their interaction, as they significantly changed their behavior when playing either against a computer opponent or in separate rooms, consistent with prior reports (Carter et al., 2012; Chang et al., 2013; Hosokawa and Watanabe, 2012). Moreover, other-predictive neurons were selectively influenced by social context compared other population cells, suggesting that these cells encoded information that was specific to other social agents rather than any information about the environment which affected outcome. The monkeys also selected the appropriate responses when their opponent's decisions were known, suggesting that they understood the consequent payoff. While the joint nature of the task precludes the possibility of identifying "involuntary errors" by the individual animals, we find that the monkeys made incorrect selections on <10% of sequential control trials making such rare occurrences highly unlikely to qualitatively affect the study's results. This conclusion is also supported by the finding that the population prediction of the opponent's decisions was robust to substantial deletion of trials. However, as with any animal or human study that investigates interactive behavior, what internal thought process truly motivates these different behaviors can only be speculated upon. On this point, we note that cooperation is based on the observable action of two interacting individuals, rather than its hidden motivation, and is defined explicitly as the selection of actions capable of leading to joint benefit but which can also lead to loss if the action is not mutual.

Taken together, the present findings support the proposed role of the dACC in encoding a dynamic model of the environment (Adolphs, 2009; Karlsson et al., 2012; Sheth et al., 2012) but considerably expand it into the inclusion of mutual interactions which require an explicit representation of another's yet unknown behavior. The two distinct groups of neurons found in the dACC, encoding the self versus predicting the other's decisions, may therefore be uniquely suitable to allow the soon-available actual decision of the opponent and known decision of the acting monkey to update the internal model of their joint decisions in a way analogous to delta-learning (Pouget and Snyder, 2000) or an actor-critic (Parush et al., 2011; Williams and Eskandar, 2006; Witten, 1977) framework. Given the broad anatomical connectivity of the dACC to areas that encode aspects of socially-guided interactions, including the temporal-parietal junction, superior temporal sulcus, amygdala and orbitofrontal cortex, the dACC is likely to be part of a wider network of areas, sometimes referred to as the "social brain." The observed role of the dACC in predicting another's intentions contributes to our under-

standing of this proposed network. For instance, disruption of its activity markedly degraded cooperative behavior, suggesting that dACC activity may be necessary for constructive interaction between individuals and social learning. Such deficits are particularly prominent in individuals with autism-spectrum disorders or antisocial behavior in which anticipating another's intentions or state of mind and incorporating them into one's actions are severely affected (Frith and Frith, 1999; Lombardo and Baron-Cohen, 2011). Our neuronal findings in combination with the behavioral effects observed with stimulation may therefore pave the way toward targeted treatment in the dACC for these or similar disorders in which dysfunctional social behavior is a predominant feature.

## EXPERIMENTAL PROCEDURES

### Task Design

Four adult male Rhesus monkeys (*Macaca Mulatta*) across four paired combinations were trained to play an iterated prisoner's dilemma (iPD) game. On successive trials, two images (an orange hexagon and a blue triangle) were randomly displayed on the left and right of the screen (Figure 1A). Each monkey selected one of the two images using a joystick and was not shown the other monkey's concurrent selection. The outcome of each monkey's selection depended on both of their concurrent choices, according to the payoff matrix shown in Figure 1B. Based on these payoffs, the orange hexagon was operationally defined as "cooperation" since mutual cooperation led to the highest mutual reward (Camerer, 2003). The blue triangle was operationally defined as "defection" since unilateral defection led to the highest individual reward. However, if both monkeys defected, they each received less reward than if they both cooperated. Note, importantly, that the terms cooperation and defection are used here solely to indicate the potential for mutual benefit or loss dependent on the opponent's selection. Mutual cooperation and mutual defection indicates that both monkeys made the same choice. See Supplemental Information for trial structure details.

### Neuronal Recording and Stimulation

#### Single-Unit Isolation and Recordings

All procedures were performed under approval by the Massachusetts General Hospital institutional review board and were conducted in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines. Prior to recordings, floating micro-electrode arrays (MicroProbes for Life Sciences) were surgically implanted in each monkey. The electrodes were implanted in the dACC through a wide craniotomy under stereotactic guidance (David Kopf Instruments). The location of the arrays was confirmed by direct visual inspection of the sulcal and gyral anatomy with the electrode tips located 8 mm from the cortical surface. Each array had 36 microelectrodes spaced horizontally 400  $\mu$ m apart. Electrode leads were secured to the skull and attached to connectors with the aid of titanium miniscrews and dental acrylic.

Recordings began 2 weeks following surgical recovery. A Plexon multi-channel acquisition processor was used to amplify and band-pass filter the neuronal signals (150 Hz–8 kHz; 1 pole low-cut and 3 pole high-cut with 1,000 $\times$  gain; Plexon). Shielded cabling carried the signals from the electrode array to a set of six 16-channel amplifiers. Neural signals were then digitized at 40 kHz and processed to extract action potentials by the Plexon workstation. Classification of the waveforms was performed using template matching and principal component analysis based on waveform parameters. Only single-, well-isolated units with identifiable waveform shapes and adequate refractory periods were used. When an individual electrode recorded more than one neuron, a high degree of isolation was required in order to include each as a single-unit ( $p < 0.01$ , multivariate ANOVA across the 1+ two principal components). We did not include multi-unit activity.

#### Electrical Stimulation Protocol

During stimulation trials, the monkeys performed the iPD and zero-sum games in separate sessions. Each session was composed of randomly selected

30–40 stimulated trials followed by another 30–40 trials in which no stimulation was delivered. Stimulation was administered as a brief series of alternating rectangular positive to negative voltage pulses. Stimulation parameters were 100  $\mu$ A and 200 Hz biphasic pulses, with cathodal phase leading. Average impedance at the time of the stimulation experiments was 100–500 k $\Omega$ . Here, all 32 electrode contacts were simultaneously stimulated per array. Stimulation was given for 1,000 ms and included the baseline and image presentation periods. Stimulation ended prior to presentation of the go cue and prior to the monkey's selection.

### Statistical Analysis

A stepwise linear regression was conducted in order to determine how the different task parameters modulated the neuronal activity. In this analysis, parameters are incrementally added to the model, starting with the parameter that explains the most variance and continuing on to the parameters that most explain the remaining variance, terminating when parameters no longer significantly explain the residual variance. The model included the four main effect parameters, as described below (self-current, other-current, self-past and other-past) as well as their pairwise interactions (see Equation 1),

$$r(t) = a + \sum_{i=1}^4 \beta_i^{\text{Main}} M_i^{\text{Main}} + \sum_{j=1}^6 \beta_j^{\text{Inter}} M_j^{\text{Inter}} \quad (\text{Equation 1})$$

where  $r(t)$  is current trial firing rate,  $M^{\text{Main}} = \{s(t), s(t-1), o(t), o(t-1)\}$  are the four main effects and  $M^{\text{Inter}} = \{s(t)s(t-1), s(t)o(t), s(t)o(t-1), s(t-1)o(t), s(t-1)o(t-1), o(t)o(t-1)\}$  are the six second order interaction terms;  $s(t)$  is current self selection,  $o(t)$  is current other selection, and  $(t-1)$  indicates preceding trial.

For brevity, “self” refers here to the selections of the monkey in which neural recordings were performed and “other” refers to the selections of the opponent (i.e., selecting to cooperate or defect). In addition, “current” refers to the two monkeys’ current selection (i.e., the trial from which neuronal activity was being evaluated) and “past” refers to the two monkeys’ selections on the previous trial. The depended variable is the averaged neuronal firing in the 500 ms period before response selection (i.e., choosing cooperation versus defection) and during the 500 ms period after selection, referred to as “pre-selection” and “post-selection,” respectively. Note that we chose to use a stepwise linear regression this analysis since the task parameters and samples were neither balanced nor independent (see further details in [Supplemental Information](#)). Multiple complimentary analyses, including a four-way analysis of variance, AIC analysis, and mixture of regressions analysis, yielded qualitatively similar results.

### SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.045>.

### ACKNOWLEDGMENTS

This project was funded by NIH 5R01-HD059852, the Presidential Early Career Award for Scientists and Engineers, and the Whitehall Foundation. Data are available as [Supplemental Information](#). We thank John Assad, Wael Asaad, Shaul Druckmann, Shaul Hochstein, Daeyeol Lee, Israel Nelken, and Sameer Sheth for insightful discussions, Caitlin Commins, Christine Emmanuel, Rebecca Gwaltney, Morgan Jamiel, and Kaitlin Sodon for technical assistance, and Katie Ris-Vicari for the graphical abstract.

Received: April 13, 2014  
Revised: October 25, 2014  
Accepted: January 5, 2015  
Published: February 26, 2015

### REFERENCES

Abe, H., and Lee, D. (2011). Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70, 731–741.

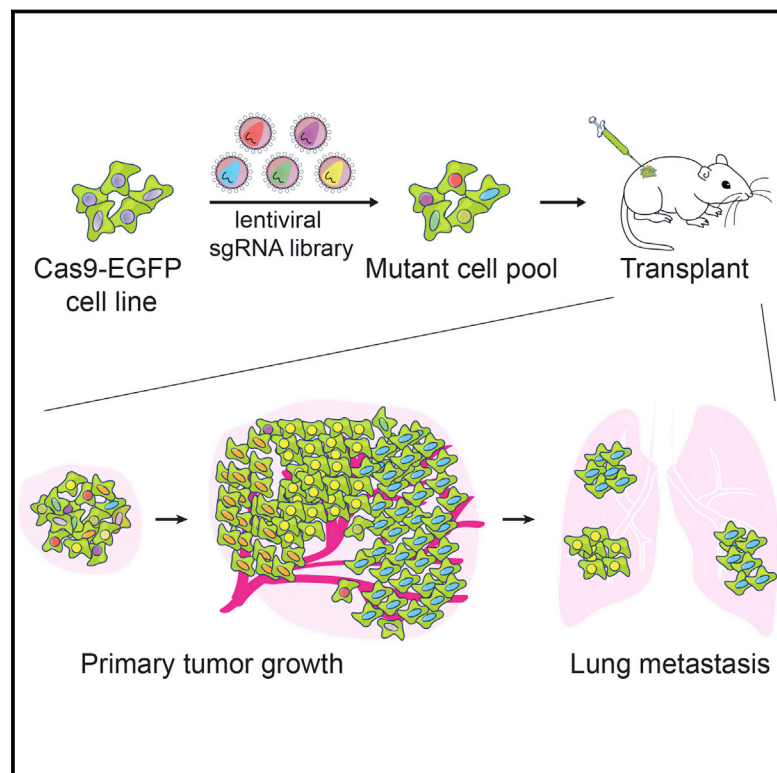
- Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716.
- Apps, M.A.J., and Ramnani, N. (2014). The anterior cingulate gyrus signals the net value of others’ rewards. *J. Neurosci.* 34, 6190–6200.
- Apps, M.A.J., Balsters, J.H., and Ramnani, N. (2012). The anterior cingulate cortex: monitoring the outcomes of others’ decisions. *Soc. Neurosci.* 7, 424–435.
- Azzi, J.C.B., Sirigu, A., and Duhamel, J.R. (2012). Modulation of value representation by social context in the primate orbitofrontal cortex. *Proc. Natl. Acad. Sci. USA* 109, 2126–2131.
- Barracough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* 7, 404–410.
- Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. *Nature* 456, 245–249.
- Behrens, T.E.J., Hunt, L.T., and Rushworth, M.F.S. (2009). The computation of social behavior. *Science* 324, 1160–1164.
- Britten, K.H., Newsome, W.T., Shadlen, M.N., Celebrini, S., and Movshon, J.A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis. Neurosci.* 13, 87–100.
- Bshary, R., Grutter, A.S., Willener, A.S.T., and Leimar, O. (2008). Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature* 455, 964–966.
- Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press).
- Carter, R.M., Bowling, D.L., Reeck, C., and Huettel, S.A. (2012). A distinct role of the temporal-parietal junction in predicting socially guided decisions. *Science* 337, 109–111.
- Chang, S.W.C., Gariépy, J.F., and Platt, M.L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci.* 16, 243–250.
- Clutton-Brock, T. (2009). Cooperation between non-kin in animal societies. *Nature* 462, 51–57.
- de Waal, F.B.M. (2000). Primates—a natural heritage of conflict resolution. *Science* 289, 586–590.
- Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8, 1611–1618.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180.
- Donahue, C.H., Seo, H., and Lee, D. (2013). Cortical signals for rewarded actions and strategic exploration. *Neuron* 80, 223–234.
- Frith, C.D., and Frith, U. (1999). Interacting minds—a biological basis. *Science* 286, 1692–1695.
- Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501.
- Hampton, A.N., Bossaerts, P., and O’Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. USA* 105, 6741–6746.
- Hosokawa, T., and Watanabe, M. (2012). Prefrontal neurons represent winning and losing during competitive video shooting games between monkeys. *J. Neurosci.* 32, 7662–7671.
- Karlsson, M.P., Tervo, D.G.R., and Karpova, A.Y. (2012). Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. *Science* 338, 135–139.
- Kuhlman, D.M., and Marshello, A.F.J. (1975). Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner’s dilemma. *J. Pers. Soc. Psychol.* 32, 922–931.
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* 11, 404–409.



- Lee, D., McGreevy, B.P., and Barraclough, D.J. (2005). Learning and decision making in monkeys during a rock-paper-scissors game. *Brain Res. Cogn. Brain Res.* 25, 416–430.
- Lombardo, M.V., and Baron-Cohen, S. (2011). The role of the self in mindblindness in autism. *Conscious. Cogn.* 20, 130–140.
- Nash, J.F. (1950). Equilibrium points in N-person games. *Proc. Natl. Acad. Sci. USA* 36, 48–49.
- Parush, N., Tishby, N., and Bergman, H. (2011). Dopaminergic balance between reward maximization and policy complexity. *Front. Syst. Neurosci.* 5, 22.
- Paus, T. (2001). Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nat. Rev. Neurosci.* 2, 417–424.
- Pouget, A., and Snyder, L.H. (2000). Computational approaches to sensorimotor transformations. *Nat. Neurosci.* 3(Suppl), 1192–1198.
- Rapoport, A., and Chammah, A.M. (1965). *Prisoner's Dilemma; A Study in Conflict and Cooperation* (Ann Arbor: University of Michigan Press).
- Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 35, 395–405.
- Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22, 1694–1703.
- Rizzolatti, G., and Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat. Rev. Neurosci.* 11, 264–274.
- Rudebeck, P.H., Buckley, M.J., Walton, M.E., and Rushworth, M.F.S. (2006). A role for the macaque anterior cingulate gyrus in social valuation. *Science* 313, 1310–1312.
- Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* 300, 1755–1758.
- Sanfey, A.G., Loewenstein, G., McClure, S.M., and Cohen, J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends Cogn. Sci.* 10, 108–116.
- Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377.
- Seo, H., Cai, X., Donahue, C.H., and Lee, D. (2014). Neural correlates of strategic reasoning during competitive games. *Science* 346, 340–343.
- Shadlen, M.N., Britten, K.H., Newsome, W.T., and Movshon, J.A. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* 16, 1486–1510.
- Sheth, S.A., Mian, M.K., Patel, S.R., Asaad, W.F., Williams, Z.M., Dougherty, D.D., Bush, G., and Eskandar, E.N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* 488, 218–221.
- Stephens, D.W., McLinn, C.M., and Stevens, J.R. (2002). Discounting and reciprocity in an iterated prisoner's dilemma. *Science* 298, 2216–2218.
- Tomlin, D., Kayali, M.A., King-Casas, B., Anen, C., Camerer, C.F., Quartz, S.R., and Montague, P.R. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050.
- Vickery, T.J., Chun, M.M., and Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* 72, 166–177.
- Vogel, K., Bussfeld, P., Newen, A., Hermann, S., Happé, F., Falkai, P., Maier, W., Shah, N.J., Fink, G.R., and Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14, 170–181.
- Warneken, F., and Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science* 311, 1301–1303.
- Williams, Z.M., and Eskandar, E.N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat. Neurosci.* 9, 562–568.
- Witten, I.H. (1977). Adaptive optimal controller for discrete-time Markov environments. *Inf. Control* 34, 286–295.
- Yoshida, K., Saito, N., Iriki, A., and Isoda, M. (2012). Social error monitoring in macaque frontal cortex. *Nat. Neurosci.* 15, 1307–1312.

# Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis

## Graphical Abstract



## Authors

Sidi Chen, Neville E. Sanjana, ...,  
Feng Zhang, Phillip A. Sharp

## Correspondence

zhang@broadinstitute.org (F.Z.),  
sharppa@mit.edu (P.A.S.)

## In Brief

Using an in vivo genome-wide CRISPR/Cas9 screen, loss-of-function mutations that drive tumor growth and metastasis to the lung have been identified, demonstrating Cas9-based screening as a robust method to systematically assay gene phenotypes in cancer evolution.

## Highlights

- Genome-wide in vivo CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis
- Screen identifies loss-of-function mutations in known tumor suppressors and novel genes
- Candidate metastasis genes are validated using a pooled competition assay
- Effect of mutations on primary tumor growth positively correlates with metastasis



# Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis

Sidi Chen,<sup>1,2,3,10</sup> Neville E. Sanjana,<sup>3,4,5,6,7,10</sup> Kaijie Zheng,<sup>3,4</sup> Ophir Shalem,<sup>3,4</sup> Kyungheon Lee,<sup>8</sup> Xi Shi,<sup>3,4</sup> David A. Scott,<sup>3,4</sup> Jun Song,<sup>8</sup> Jen Q. Pan,<sup>3,4</sup> Ralph Weissleder,<sup>8,9</sup> Hakho Lee,<sup>8</sup> Feng Zhang,<sup>3,4,5,6,7,\*</sup> and Phillip A. Sharp<sup>1,2,\*</sup>

<sup>1</sup>David H. Koch Institute for Integrative Cancer Research, Cambridge, MA 02139, USA

<sup>2</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

<sup>4</sup>Stanley Center for Psychiatric Research, Cambridge, MA 02142, USA

<sup>5</sup>McGovern Institute for Brain Research

<sup>6</sup>Department of Brain and Cognitive Sciences

<sup>7</sup>Department of Biological Engineering

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>8</sup>Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114, USA

<sup>9</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>10</sup>Co-first author

\*Correspondence: [zhang@broadinstitute.org](mailto:zhang@broadinstitute.org) (F.Z.), [sharp@mit.edu](mailto:sharp@mit.edu) (P.A.S.)

<http://dx.doi.org/10.1016/j.cell.2015.02.038>

## SUMMARY

Genetic screens are powerful tools for identifying genes responsible for diverse phenotypes. Here we describe a genome-wide CRISPR/Cas9-mediated loss-of-function screen in tumor growth and metastasis. We mutagenized a non-metastatic mouse cancer cell line using a genome-scale library with 67,405 single-guide RNAs (sgRNAs). The mutant cell pool rapidly generates metastases when transplanted into immunocompromised mice. Enriched sgRNAs in lung metastases and late-stage primary tumors were found to target a small set of genes, suggesting that specific loss-of-function mutations drive tumor growth and metastasis. Individual sgRNAs and a small pool of 624 sgRNAs targeting the top-scoring genes from the primary screen dramatically accelerate metastasis. In all of these experiments, the effect of mutations on primary tumor growth positively correlates with the development of metastases. Our study demonstrates Cas9-based screening as a robust method to systematically assay gene phenotypes in cancer evolution in vivo.

## INTRODUCTION

Cancer genomes have complex landscapes of mutations and diverse types of genetic aberrations (Lawrence et al., 2013; Weinberg, 2007). A major challenge in understanding the cancer genome is to disentangle alterations that are driving the processes of tumor evolution from passenger mutations (Garraway and Lander, 2013). Primary tumor growth and metastasis are distinct yet linked processes in the progression of solid tumors (Nguyen et al., 2009; Valastyan and Weinberg, 2011; Vanharanta and Massagué, 2013). It has been observed in the clinic that the

probability of detecting metastases in a patient correlates positively with the size of a primary tumor (Heimann and Hellman, 1998). Several possible explanations have been suggested: metastatic properties may only be acquired in late-stage tumors, larger tumors may seed proportionally more cells into circulation that eventually migrate to other sites, or cells with a strong ability to proliferate may also have enhanced ability to metastasize (Weinberg, 2007). In early studies using random insertional mutagenesis, it was observed that metastatic cell subpopulations overgrow to complete dominance in the primary tumor, suggesting progressive selection at both sites (Korczak et al., 1988; Waggoner et al., 1988).

Genetic screens are powerful tools for assaying phenotypes and identifying causal genes in various hallmarks of cancer progression (Hanahan and Weinberg, 2011). RNAi and overexpression of open reading frames (ORFs) have been utilized for screening cancer genes in several models of oncogenesis in mice (Schramek et al., 2014; Shao et al., 2014; Zender et al., 2008). Recently, the Cas9 nuclease (Barrangou et al., 2007; Bolotin et al., 2005; Chylinski et al., 2013, 2014; Deltcheva et al., 2011; Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012; Sapranaukas et al., 2011) from the microbial type II CRISPR (clustered regularly interspaced short palindromic repeats) system has been harnessed to facilitate loss-of-function mutations in eukaryotic cells (Cong et al., 2013; Mali et al., 2013). When the Cas9 nuclease is targeted to specific locations in the genome, DNA cleavage results in double-stranded breaks (DSBs), which are repaired via non-homologous end-joining (NHEJ) (Rouet et al., 1994). NHEJ repair results in insertion or deletion (indel) mutations that can cause loss of function if the DSB occurs in a coding exon. The Cas9 nuclease can be guided to its DNA target by a single-guide RNA (sgRNA) (Jinek et al., 2012), a synthetic fusion between the CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) (Deltcheva et al., 2011). In cells, Cas9-mediated gene disruption requires the full-length tracrRNA (Cong et al., 2013; Mali et al., 2013), in which secondary structures at the 3' end of tracrRNA are

critical for Cas9-mediated genome modification (Cong et al., 2013; Hsu et al., 2013).

Screens utilizing Cas9 have identified genes that are essential for cell survival and genes involved in drug resistance in various cell lines (Shalem et al., 2014; Wang et al., 2014; Koike-Yusa et al., 2014; Zhou et al., 2014). In vivo pooled screens are challenging due to many factors, such as the complexity of the library, limitations of virus delivery and/or cell transplantation, uniformity of viral transduction at a low MOI, and the complex dynamics and interactions of cells in animals. In this study, we report a genome-wide Cas9 knockout screen in a mouse model of tumor evolution. This screen provides a systematic phenotypic measurement of loss-of-function mutations in primary tumor growth and metastasis.

## RESULTS

### CRISPR/Cas9 Library-Mediated Mutagenesis Promotes Metastasis

We derived and cloned a cell line (Chen et al., 2014) from a mouse non-small-cell lung cancer (NSCLC) (Kumar et al., 2009). This cell line possesses an oncogenic *Kras* in conjunction with homozygous *p53* and heterozygous *Dicer1* loss of function (*Kras*<sup>G12D/+</sup>; *p53*<sup>-/-</sup>; *Dicer1*<sup>+/-</sup>, denoted KPD) and is capable of inducing tumors when transplanted into immunocompromised mice (Chen et al., 2014; Kumar et al., 2009). We transduced this cell line with a lentivirus carrying a Cas9 transgene fused to a GFP and generated clonal cell lines (Cas9-GFP KPD) (Experimental Procedures) (Figures S1A and S1B). A clonal Cas9-GFP KPD cell line (clone 5) was selected to provide genetic and cellular homogeneity for subsequent screens.

We utilized a pooled genome-wide mouse sgRNA library (termed mouse genome-scale CRISPR knockout library A, or mGeCKOa) containing 67,405 sgRNAs targeting 20,611 protein-coding genes and 1,175 microRNA precursors in the mouse genome (Sanjana et al., 2014). The library also contains 1,000 control sgRNAs (termed non-targeting sgRNAs) designed to have minimal homology to sequences in the mouse genome (Sanjana et al., 2014; Shalem et al., 2014). We transduced the Cas9-GFP KPD cell line with the mGeCKOa library in three independent infection replicate experiments; for each replicate, the library representation (cells per lentiviral CRISPR construct) was greater than 400× (Figure 1A) (Experimental Procedures).

After in vitro culture for 1 week, we subcutaneously transplanted  $3 \times 10^7$  cells into the flanks of immunocompromised *Nu/Nu* mice (Figure 1A). We transplanted the cells from each infection replicate into four mice, using one mouse for early tumor sequencing and three mice for sequencing of late-stage primary tumor and metastases (Figure 1A). Both mGeCKOa-transduced and untransduced Cas9-GFP KPD cells formed tumors at the injection site (Figure 1B). Like most subcutaneously transplanted tumors, these tumors were poorly differentiated. The primary tumors induced by mGeCKOa-transduced cells grew slightly faster than tumors from the untransduced cells at an early stage (Figure 1C) (2 weeks post-transplantation) (paired two-tailed t test,  $p = 0.05$ ), but at late stages all tumors were similar in size (paired two-tailed t test,  $p = 0.18$  for data at 4 weeks,  $p = 0.6$  for data at 6 weeks) (Figure 1C).

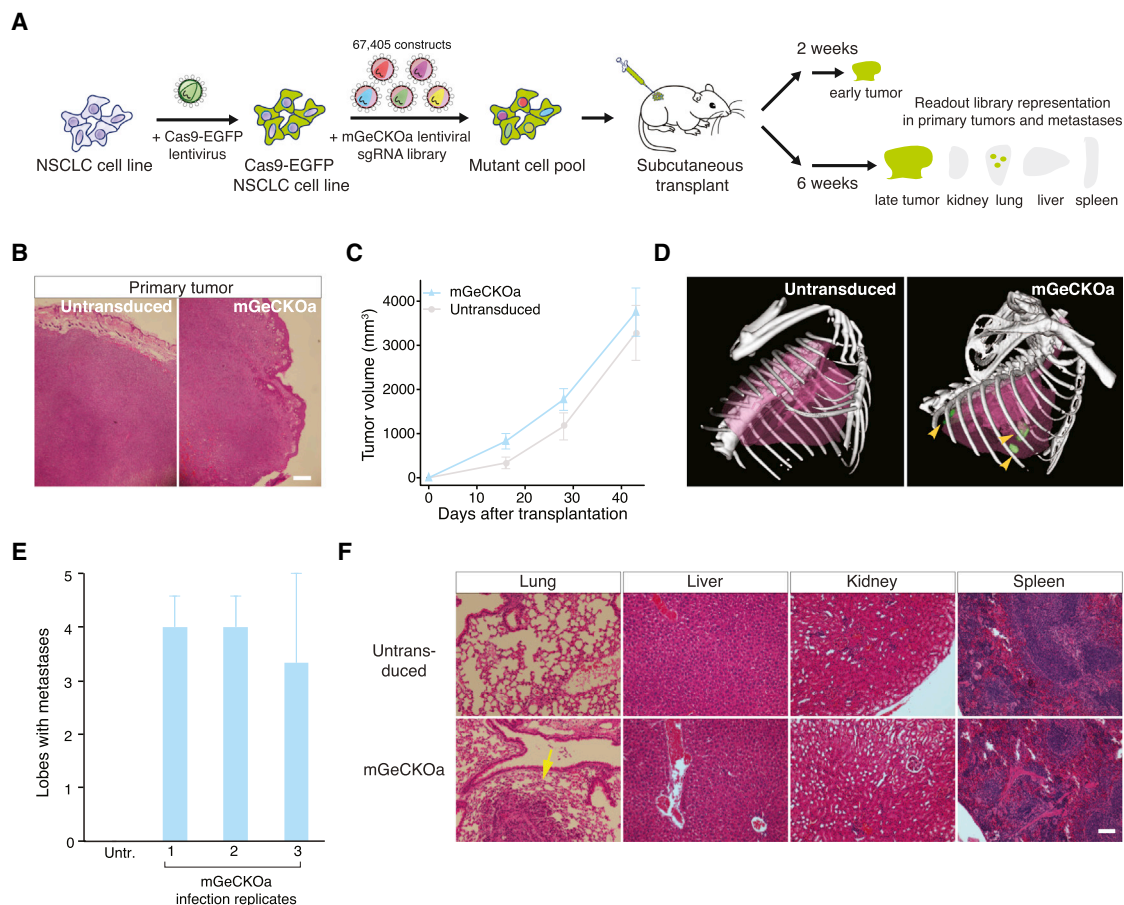
At 6 weeks post-transplantation, we imaged the mice using micro-computed tomography ( $\mu$ CT) and found tumors in the lungs of the mice transplanted with mGeCKOa-transduced Cas9-GFP KPD cells (mGeCKOa mice), but not in the mice transplanted with untransduced Cas9-GFP KPD cells (control mice) (Figure 1D, Figure S1C). Mice were sacrificed and examined for metastases in various organs. Under a fluorescent stereoscope at 6× magnification, metastases were visually detected in the lung in 89% (8/9) of the mGeCKOa mice (Figure S1D). The mGeCKOa mice on average had 80% of their lung lobes positive for metastases (Figure 1E). In contrast, none (0/3) of the control mice developed detectable metastases in the lung (Figure 1E). At this time, metastases were not detected in the liver, kidney, or spleen in either group (Figure 1F). These data indicated that mGeCKOa library transduction enhanced the ability of the Cas9-GFP KPD cells to form metastases in the lung.

### Dynamic Evolution of sgRNA Library Representation during Tumor Growth and Metastasis

To investigate the sgRNA representation through different stages of tumor evolution and to identify genes where loss of function confers a proliferative or metastatic phenotype, we used deep sequencing to readout the sgRNA representation (see Data S1 in Dataset S1). At 6 weeks post transplantation, we sequenced the late-stage primary tumor and three random lobes from the lung of each of the nine mGeCKOa mice (Figure 1A) (Experimental Procedures). In parallel, we also sequenced the mGeCKOa input plasmid library, the pre-transplantation mGeCKOa-transduced Cas9-GFP KPD cells (cultured in vitro for 7 days after transduction), and early-stage primary tumors (2 weeks post transplantation, one mouse from each infection replicate). In the cell samples, the sgRNA representations showed high concordance between technical replicates (correlation,  $\rho = 0.95$  on average,  $n = 3$ ) and biological infection replicates (correlation,  $\rho = 0.84$  on average,  $n = 3$ ) (Figures 2A, S2A, S2B, and S2E). The sgRNA representation of cell samples correlates highly with the plasmid representation (correlation,  $\rho = 0.93$  on average,  $n = 3$ ) (Figures 2A, S2C, and S2E). Furthermore, different sgRNAs that target the same gene are correlated in terms of rank change (correlation,  $\rho = 0.49$  on average,  $n = 3$ ) (Figure S2D). Using gene set enrichment analysis (GSEA), we found that the sgRNAs with significantly decreased abundance in cells compared to plasmid are enriched for genes involved in fundamental cellular processes, such as ribosomal proteins, translation factors, RNA splicing factors, and RNA processing factors, indicating selection against the loss of these genes after 1 week in culture (Figure S2F).

To investigate the sgRNA library dynamics in different sample types (plasmid, pre-transplantation cells, early primary tumor, late primary tumor, and lung metastases), we compared the overall distributions of sgRNAs from all samples sequenced. Cell samples clustered tightly with each other and the plasmid, forming a cell-plasmid clade (Figures 2A and S2E). Early primary tumor samples also clustered with each other and then with the cell-plasmid clade, whereas late tumors and lung metastases clustered together in a distinct group (Figures 2A and S2E). The overlap of detected sgRNAs between different pre-transplantation infection replicates is over 95% (Figure S3A). The detected sgRNAs in the three infection replicates of early tumor





**Figure 1. Tumor Growth and Metastasis in Transplanted Cas9-GFP KPD Cells with mGeCKOa Library**

(A) Schematic representation of the loss-of-function metastasis screen using the mouse genome-scale CRISPR/Cas9 knockout library (mGeCKOa).

(B) Representative H&E stains of primary tumor from *Nu/Nu* mice subcutaneously transplanted with a Cas9-GFP *Kras*<sup>G12D/+</sup>; *p53*<sup>-/-</sup>; *Dicer1*<sup>+/-</sup> (KPD) NSCLC cell line that was either untransduced or transduced with the mGeCKOa lentiviral library. Scale bar, 200  $\mu$ m.

(C) Primary tumor growth curve of *Nu/Nu* mice transplanted with untransduced cells (*n* = 3 mice) or mGeCKOa-transduced Cas9-GFP KPD cells (*n* = 9 mice). Error bars indicate SEM.

(D) MicroCT 3D reconstruction of the lungs of representative mice transplanted with control (untransduced) and mGeCKOa-transduced (mGeCKOa) cell pools. Lung metastases were identified and traced in each 2D section (green).

(E) Percent of lobes with metastases visible after dissection under a fluorescence stereoscope in *Nu/Nu* mice transplanted with untransduced Cas9-GFP KPD cells (*n* = 3 mice) or mGeCKOa-transduced Cas9-GFP KPD cells with three independent infection replicate experiments (1, 2, and 3; *n* = 3 mice per replicate). Error bars indicate SEM.

(F) Representative H&E stains from various organs of *Nu/Nu* mice subcutaneously transplanted with untransduced and mGeCKOa-transduced Cas9-GFP KPD cells. Yellow arrow indicates a lung metastasis. Scale bar, 40  $\mu$ m.

See also Figure S1.

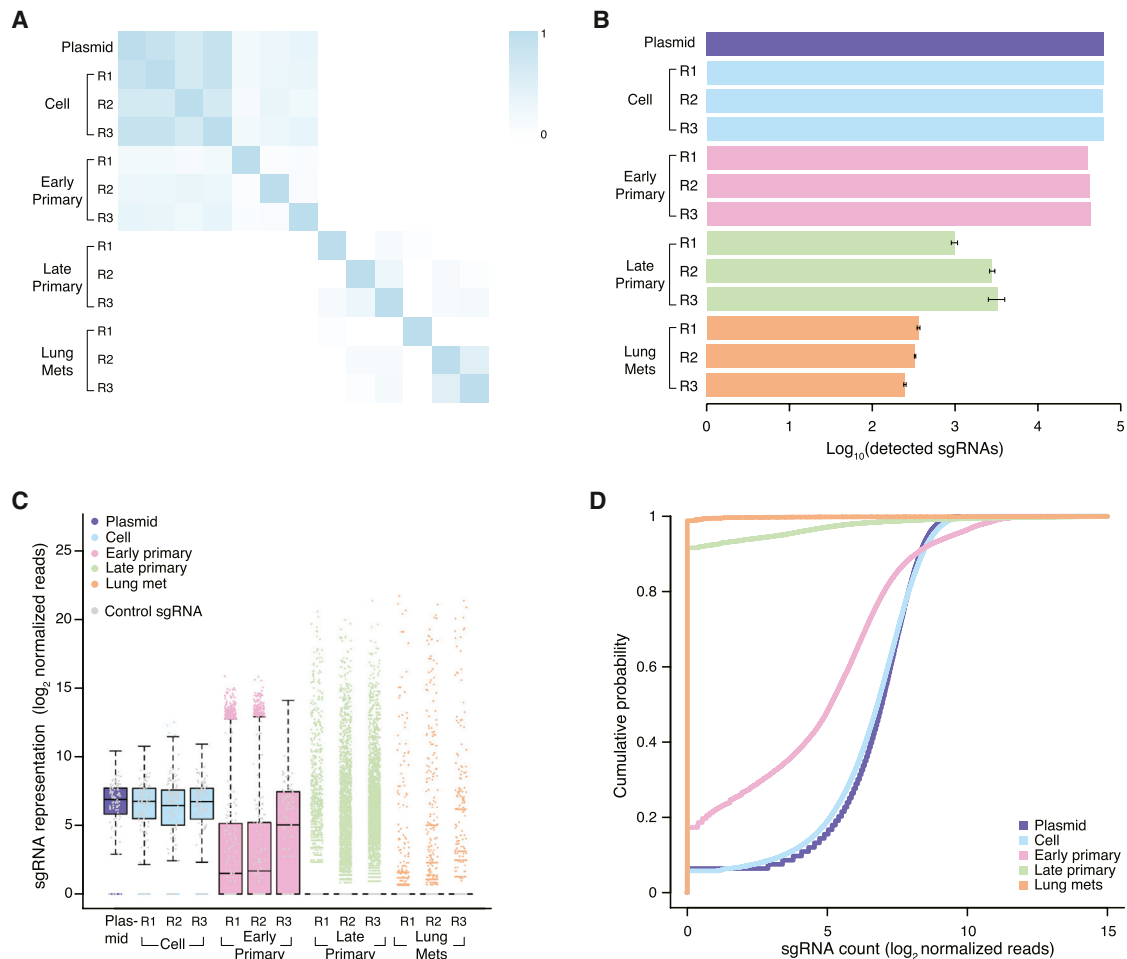
samples overlap 63%–76% with each other (Figure S3B). Early primary tumors retained less than half (32%–49%) of the sgRNAs found in the transplanted cell populations (Figures 2B, 2C, S3C, and S3D). Compared to the cell populations, sgRNAs whose targets are genes involved in fundamental cellular processes are further depleted in early tumors (Table S1).

Interestingly, only a small fraction of sgRNAs (less than 4% of all sgRNAs, or less than 8% of sgRNAs in the early primary tumor of the corresponding replicate) were detected in the late-stage primary tumor samples (Figures 2B, 2C, S3C, and S3D). The sgRNA diversity (i.e., number of different sgRNAs detected) further decreased in samples from lung metastases (Figures

2B, 2C, S3C, and S3D). The lung samples retained  $\leq 0.4\%$  of all sgRNAs in the mGeCKOa library, or  $\leq 1.1\%$  of sgRNAs found in the early primary tumor of the corresponding replicate, with a subset of highly enriched sgRNAs (Figures 2B, 2C, S3C, and S3D). The global patterns of sgRNA distributions in different sample types are distinct, as is evident in the strong shifts in the respective cumulative distribution functions (Kolmogorov-Smirnov [KS] test,  $p < 10^{-15}$  for all pairwise comparisons) (Figure 2D).

#### Enriched sgRNAs in Primary Tumors

Late primary tumors retain few sgRNAs (on average  $813 \pm 264$  sgRNAs, *n* = 9 mice), with even fewer at high frequencies



**Figure 2. Representation of mGeCKOa Library at Different Stages of Tumor Growth and Metastasis**

(A) Pearson correlation coefficient of the normalized sgRNA read counts from the mGeCKOa plasmid library, transduced cells before transplantation (day 7 after spinfection), early primary tumors (~2 weeks after transplantation), late primary tumors (~6 weeks after transplantation), and lung metastases (~6 weeks after transplantation). For each biological sample type, three independent infection replicates (R1, R2, and R3) are shown.  $n = 1$  mouse per infection replicate for early primary tumors;  $n = 3$  mice per infection replicate for late primary tumors and lung samples.

(B) Number of unique sgRNAs in the plasmid, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Error bars for late primary tumors and lung metastases denote SEM for  $n = 3$  mice per infection replicate.

(C) Boxplot of the sgRNA normalized read counts for the mGeCKOa plasmid pool, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Outliers are shown as colored dots for each respective sample. Gray dots overlaid on each boxplot indicate read counts for the 1,000 control (non-targeting) sgRNAs in the mGeCKOa library. Distributions for late primary tumors and lung metastases are averaged across individual mice from the same infection replication.

(D) Cumulative probability distribution of library sgRNAs in the plasmid, cells before transplantation, early and late primary tumors, and lung metastases as in (A). Distributions for each sample type are averaged across individual mice and infection replications.

See also Figures S2 and S3.

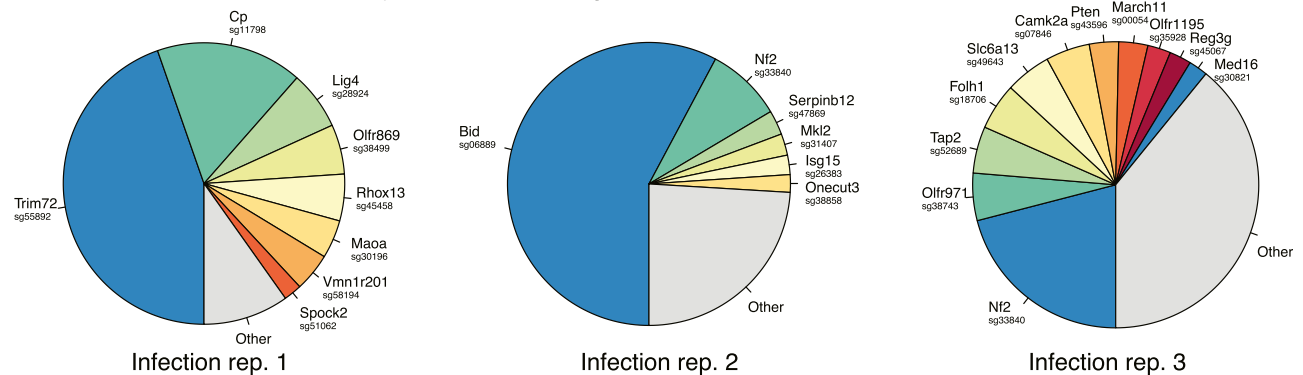
( $4 \pm 1$  sgRNAs with  $>5\%$  of total reads) in each mouse (Figures 2B, 2C, S2C, S2D, 3A, and S4H). We used three methods to identify enriched sgRNAs in late primary tumors: (1) sgRNAs above a certain threshold, (2) top-ranked sgRNAs in the tumor of each mouse, and (3) using false discovery rate (FDR), i.e., sgRNAs enriched compared to the distribution of the 1,000 non-targeting sgRNAs. All three methods generated similar results (Figure S4A). Taking the results from (3) as an example, a total of 935 sgRNAs (targeting 909 genes) are enriched over the non-targeting controls (FDR cutoff = 0.2%) in the late primary

tumor of one or more mice (Figures 3B and 3C). These sgRNAs are targeting genes highly enriched in apoptosis pathways (Table S2), with many of them being pro-apoptotic, such as BH3 interacting-domain death agonist (*Bid*), phosphatase and tensin homolog (*Pten*), cyclin-dependent kinase inhibitor 2a (*Cdkn2a*), and O-6-methylguanine-DNA methyltransferase (*Mgmt*), suggesting strong selection for mutations that inactivate apoptosis in primary tumor cells.

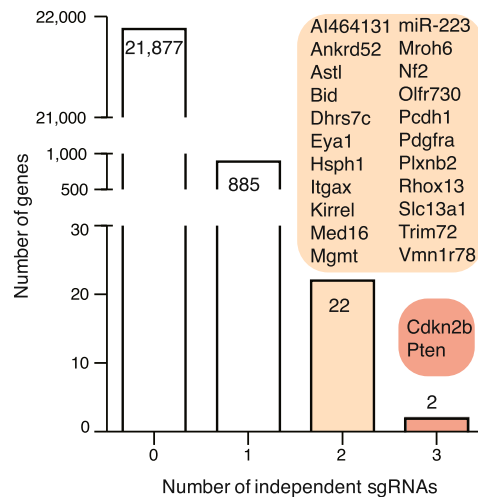
We identified 24 candidate genes that were targeted by two or more independent sgRNAs enriched in late primary tumors

A

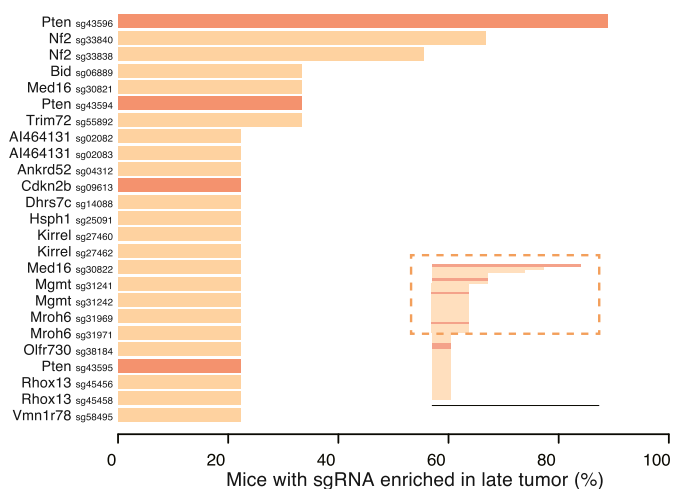
## Primary tumor (6 weeks) sgRNA distributions from individual mice



B



C



## Figure 3. Enriched sgRNAs from the mGeCKOa Screen in Primary Tumors

(A) Pie charts of the most abundant sgRNAs in the primary tumors (at ~6 weeks post-transplantation) of three representative mice (one from each replicate mGeCKOa infection). The area for each sgRNA corresponds to the fraction of total reads from the primary tumor for the sgRNA. All sgRNAs with  $\geq 2\%$  of total reads are plotted individually.

(B) Number of genes with 0, 1, 2, or 3 significantly enriched (FDR  $< 0.2\%$  for at least one mouse) mGeCKOa sgRNAs targeting that gene. For genes/miRs with 2 or more enriched sgRNAs, genes/miRs are categorized by how many sgRNAs targeting that gene/miR are enriched as indicated in the colored bubbles adjacent to each bar.

(C) Inset: waterfall plot of sgRNAs where multiple sgRNAs targeting the same gene are significantly enriched in primary tumors. Each sgRNA is ranked by the percent of mice in which it is enriched. Only sgRNAs enriched in two or more mice are shown in the main panel. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

See also Figures S3, S4, and S5.

(Figures 3B and 3C). These genes were found to be mutated in patients in many previously reported cancer sequencing studies curated by cBioPortal (Cerami et al., 2012; Gao et al., 2013) (Figure S5A). For example, in somatic mutations identified by The Cancer Genome Atlas (TCGA) for NSCLC, including adenocarcinoma (LUAD) (Cancer Genome Atlas Research Network, 2014) and lung squamous cell carcinoma (LUSC) (Cancer Genome Atlas Research Network, 2012), 36% (107/407) of patients have one or more of these 24 genes mutated (Figures S5B and S5C). Several candidates were well-known tumor suppressors,

such as *Pten*, cyclin-dependent kinase inhibitor 2b (*Cdkn2b*), neurofibromin 2 (*Nf2/Merlin*), alpha-type platelet-derived growth factor receptor (*Pdgfra*), and integrin alpha X (*Itgax*).

## Enriched sgRNAs in Metastases

We also sequenced the sgRNA distributions from three lung lobes for each mouse transplanted with mGeCKOa-transduced Cas9-GFP KPD cells. In each lobe, the sgRNA representation is dominated by one or a few sgRNAs (Figures 4A, S3D, and S4I). In each mouse, the lung sgRNA representation (average of

normalized sgRNA representations from three lobes) is also dominated by a small number of sgRNAs (on average,  $3.4 \pm 0.4$  sgRNAs with  $>5\%$  of total reads) (Figure 4B), suggesting that metastases were seeded by a small set of cells, which grew to dominance over this timescale. Non-targeting sgRNAs were occasionally detected in the metastases but were never observed at high frequency ( $<0.1\%$  of total reads in any lobe; Figures 2C, 4A and 4B, and S4I). These observations are consistent with our finding that untransduced tumors are not metastatic (Figure 1E), suggesting that specific sgRNA-mediated mutations led to metastasis.

The sgRNA representations in the lung metastases are similar to those in the late-stage primary tumors in several ways. First, the detected sgRNAs in lung samples overlap significantly with those in late tumor samples (chi-square test,  $p < 10^{-15}$ ) (Figure S3E). Second, the number of sgRNAs detected in lung samples correlates, albeit weakly, with the number of sgRNAs detected in late primary tumor samples ( $\rho = 0.42$ , F test,  $p = 0.097$ ) (Figure S3F). Third, the abundance (number of reads) of sgRNAs in the lung correlates positively with that in the late primary tumors of the same mouse (correlation,  $\rho = 0.18$  on average, F test,  $p < 0.01$ ,  $n = 9$ ) (Figure S3G). Fourth, in most mice (8/9), the lung metastasis enriched sgRNAs also occupy a large fraction of reads in the late primary tumor of the same mouse (Figure 4C, left panel), significantly larger than a random sampling of the same number of sgRNAs from the mGeCKOa library (Figure 4C, right panel). These data indicate that mutants with preferential ability to proliferate in late primary tumors are more likely to dominate the metastases.

The three methods (threshold, rank, or FDR) of finding enriched sgRNAs in the lung metastases yield similar results (Figure S4B). Using the non-targeting sgRNA distribution to set a FDR-based cutoff for enrichment, the enriched sgRNAs in different lobes of the same mouse overlap with each other by  $62\% \pm 5\%$  (chi-square test,  $p < 10^{-15}$ ) (Figure S4C), while different mice show greater variability while still overlapping significantly ( $29\% \pm 3\%$ , chi-square test,  $p < 10^{-15}$ ) (Figure S4D). The overlap between sgRNAs in different biological/infection replicate experiments when pooling enriched sgRNAs from all mice in the same replicate is  $54\%$  (chi-square test,  $p < 10^{-15}$ ) (Figure S4E), suggesting that pooling sgRNAs from mice in the same experiment facilitates the identification of shared hits. These data suggest that the three independent experiments reproducibly captured a common set of hits and provide a picture for in vivo experimental variation between different lobes, different animals, and different infection replicates.

We found 147 sgRNAs enriched in more than one lobe, and 105 sgRNAs enriched in the lung of more than one mouse (Figures 4D and 4E). These include sgRNAs targeting *Nf2*, *Pten*, tripartite motif-containing protein 72 (*Trim72*), fibrinogen alpha chain (*Fga*), *Bid*, cyclin-dependent kinase inhibitor 2a (*Cdkn2a*), zinc finger FYVE domain-containing 28 (*Zfyve28*), reproductive homeobox 13 (*Rhox13*), and BRISC and BRCA1 A complex member 1 (*Babam1*), as well as microRNA genes *miR-152* and *miR-345*. Intriguingly, a few sgRNAs targeting the Pol II subunits and olfactory receptor are also enriched in the lung, possibly due to off-target effects or unknown roles of these genes. For most sgRNAs detected in lung metastases, the relative abundance in metastases is lower than that in the late primary tumor of the

same mouse, with a metastasis-primary ratio (MPR) less than 1 (Figure S4F), likely due to more skewed distributions of sgRNAs in the metastases compared to those in the late primary tumors. A small subset of sgRNAs, however, are more abundant in metastases than in primary tumors (MPR  $> 1$ ) in multiple mice, e.g., sgRNAs targeting *Nf2*, *Trim72*, prostaglandin E synthase 2 (*Ptges2*), or ubiquitin-conjugating enzyme E2G 2 (*Ube2g2*) (Figure 4F).

For four genes, *Nf2*, *Pten*, *Trim72*, and *Zfyve28*, two independent sgRNAs targeting different regions of the same gene were enriched in lung metastases (Figure 4G). One of the *Zfyve28*-targeting sgRNAs, however, is enriched in only one mouse, whereas *Nf2*, *Pten*, and *Trim72* all have two sgRNAs enriched in multiple mice (Figure 4H). These three genes, several representative genes with one frequently enriched sgRNA (*Cdkn2a*, *Fga*, and *Cryba4*), and the two top-scoring microRNAs (*miR-152* and *miR-345*) were chosen to assay individually for primary tumor growth and metastases formation.

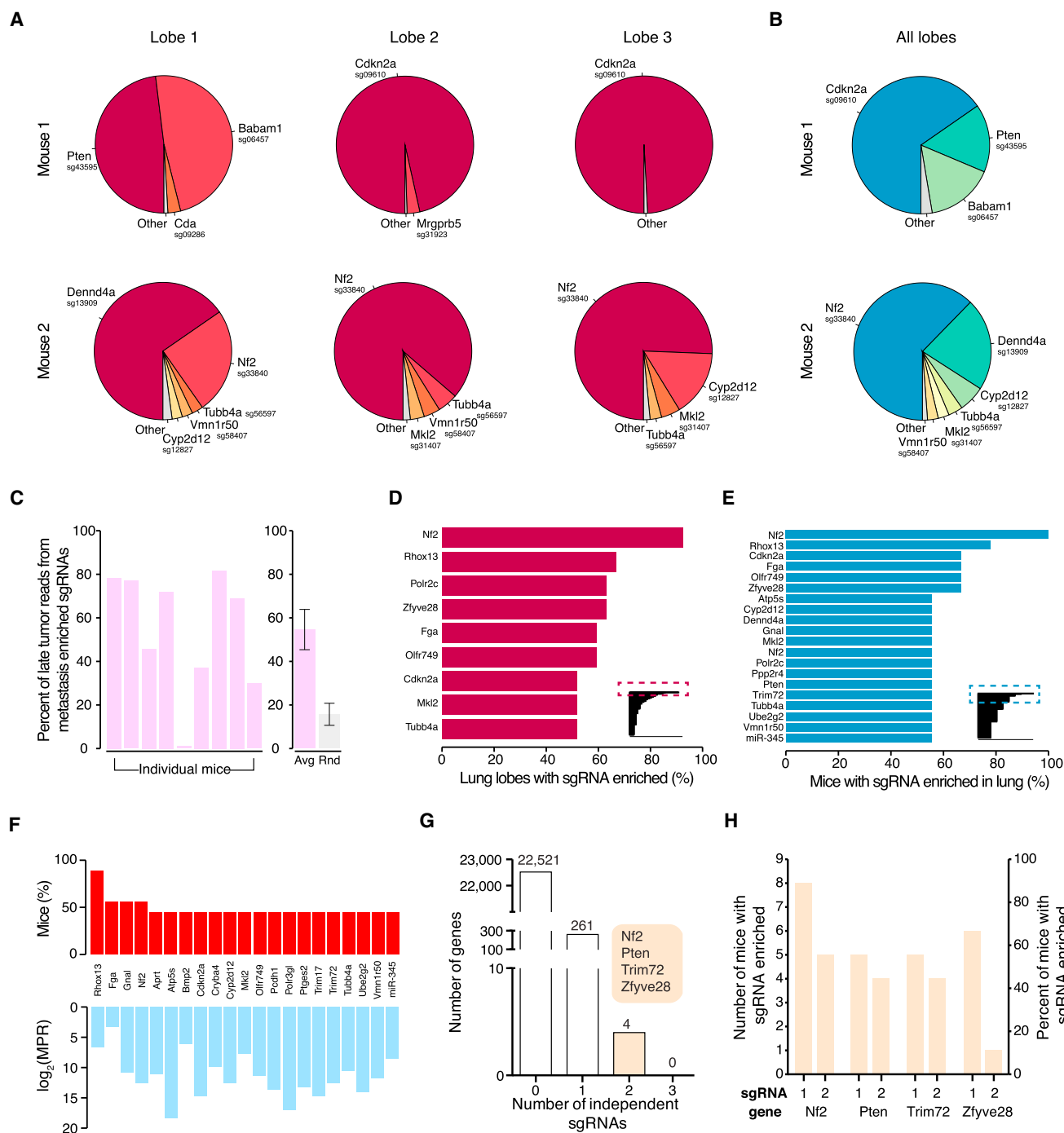
### Validation In Vivo Using Individual sgRNAs

For these eight genes (*Nf2*, *Pten*, *Trim72*, *Cdkn2a*, *Fga*, *Cryba4*, *miR-152*, and *miR-345*), we cloned multiple sgRNAs targeting each of them into the lentiGuide-Puro vector and transduced them into the Cas9-GFP KPD cell line (Figure 5A) (Experimental Procedures). As expected, these sgRNAs generated a broad distribution of NHEJ-mediated indels at the target site when examined 3 days post-transduction, with a bias toward deletions (Figure 5B). For protein-coding genes, the majority ( $>80\%$ ) of indels are out of frame, which potentially disrupts the protein functions. For *miR-152* and *miR-345*, the sgRNAs generated mostly deletions ( $>90\%$  of indels are deletions, average indel size  $\sim 7$  bp) (Figure 5B), overlapping with the loop or mature microRNA sequences in the hairpins, which are structures required for maturation of microRNAs. For proteins where specific antibodies are available (*Nf2* and *Pten*), we found that the majority of the protein products were significantly reduced 1 week after lentiviral sgRNA infection (Figure S6A).

When these single-sgRNA-transduced cells were transplanted into the flanks of immunocompromised mice, they all formed tumors in situ. With two mice injected per sgRNA and three sgRNAs per gene, all genes tested showed increased lung metastasis formation compared to controls (untransduced and non-targeting sgRNAs), with the most significant ones being *Nf2*, *Pten*, and *Cdkn2a* (Fisher's exact test, one-tailed,  $p < 10^{-3}$ ) (Figures 5C and 5D). *Fga* and *Trim72* also have effects on metastasis acceleration (*Fga*  $p = 0.001$ , *Trim72*  $p = 0.046$ ). *Cryba4* is not statistically different from controls ( $p = 0.1$ ). sgRNAs targeting *miR-345* or *miR-152* significantly increased the rate of metastasis (*miR-345*  $p = 0.01$ , *miR-152*  $p = 0.046$ ). These data suggest that loss-of-function mutations in any of *Nf2*, *Pten*, *Cdkn2a*, *Trim72*, *Fga*, *miR345*, or *miR-152* are sufficient to accelerate the rate of metastasis formation in this genetic background.

Most genes targeted by single sgRNAs also contributed to accelerated primary tumor growth compared to controls (Figure 5E). *Nf2* and *Pten* loss of function dramatically speed up tumor growth (KS test,  $p < 0.001$ ) (Figure 5E); *Cdkn2a*-, *Trim72*-, and *Fga*-targeting sgRNAs slightly accelerate primary tumor growth (KS test,  $p = 0.003$ – $0.01$ ); *Cryba4* has a marginal effect





**Figure 4. Enriched sgRNAs from the mGeCKOa Screen in Lung Metastases**

(A) Pie charts of the most abundant sgRNAs in three individual lobes of the lungs of two representative mice transplanted with mGeCKOa-transduced cells. The area for each sgRNA corresponds to the fraction of total reads from the lobe for the sgRNA. All sgRNAs with  $\geq 2\%$  of total reads are plotted individually.

(B) Pie charts of the most abundant sgRNAs in the lung (averaged across three individual lobes) for the two mice shown in (A). All sgRNAs with  $\geq 2\%$  of average reads are plotted individually.

(C) Left: percentage of late tumor reads for the significantly enriched ( $FDR < 0.2\%$ ) mGeCKOa sgRNAs found in the lung metastases (averaged across three dissected lobes). Right: in purple, the percentage of late tumor reads for the significantly enriched ( $FDR < 0.2\%$ ) mGeCKOa sgRNAs found in the lung metastases (average across all mice,  $n = 9$  mice). In gray, the percentage of late tumor reads for random, size-matched samplings of sgRNAs present in the late tumor ( $n = 100$  samplings). Error bars indicate SD.

(legend continued on next page)

(KS test,  $p = 0.08$ ); and neither *miR-152*- nor *miR-345*-targeting sgRNAs promote primary tumor growth (KS test,  $p > 0.1$ ). Overall, for the targets we examined using individual sgRNAs, the number of lobes with lung metastases strongly correlates with the terminal volume of the late primary tumor (or average primary tumor growth rate) (correlation,  $\rho = 0.83$ , F test,  $p < 0.01$ ) (Figure 5F), indicating at a single-gene level that mutant cells with a stronger ability to promote primary tumor growth generate metastases faster.

To analyze blood samples for the presence of circulating tumor cells (CTCs), we designed a microfluidic device based on the physical size of the Cas9-GFP KPD cells (Figures S6B and S6C). We performed CTC capture with terminal blood samples from mice injected with Cas9-GFP KPD cells transduced with sgRNAs targeting *Nf2*, *Pten*, *Trim72*, *Cdkn2a*, and *miR-152* and from mice injected with Cas9-GFP KPD control cells (untransduced or non-targeting sgRNA) (Figures S6C and S6D). Mice transplanted with cells transduced with sgRNAs targeting *Nf2*, *Pten*, *Trim72*, or *Cdkn2a* had a higher concentration of CTCs as compared to controls (Figures S6D–S6G), consistent with the higher rate of lung metastasis formation.

### Competitive Dynamics of Top Hits Assessed Using an sgRNA Minipool

To better understand the relative metastatic potential of multiple genes from our genome-wide screen, we designed a targeted pooled screen with a smaller library. This small library (termed validation minipool) contains 524 sgRNAs targeting 53 genes that had highly enriched sgRNAs in lung metastases in the genome-wide screen (ten sgRNAs per gene for most genes) plus 100 non-targeting sgRNAs. We also created a size-matched library containing 624 non-targeting sgRNAs (termed control minipool) (Figure 6A). Lentiviruses from these two pools were used to transduce the Cas9-GFP KPD cells, which were cultured in vitro for 1 week and then transplanted into *Nu/Nu* mice (Figure 6A). Both validation minipool- and control minipool-transduced cells induced primary tumor growth at a similar rate (Figure 6B). However, mice transplanted with validation minipool cells had a dramatically elevated rate of lung metastasis formation (Figure 6C).

We sequenced the validation minipool plasmid library and the transduced cells pre-transplantation, as well as the late-stage primary tumors and whole lungs of the mice at 5 weeks post-transplantation (see Data S2 in Dataset S1). The sgRNA representations correlate strongly between technical replicates of

the transduced cell pool, late primary tumors, and lung metastases (Figures S7A and S7D). The sgRNA representation in the cell sample strongly correlated with the plasmid (correlation,  $\rho = 0.91$ ) (Figures S7B and S7D). Almost all (99.4%) sgRNAs were recovered in the plasmid and the cell population (Figure S7C). The late primary tumors retained less than half of the sgRNAs, and the metastases in the whole lung retained only a small fraction (2%–7%) of all sgRNAs (Figure S7C). Enriched sgRNAs from lung metastases clustered with each other and with late primary tumors (Figure S7D). Similar to the genome-wide library, in this validation minipool, the plasmid and cell samples had a unimodal distribution of sgRNAs, whereas the late primary tumors and lung metastases contained a bimodal distribution, with the majority of sgRNAs being absent and a small fraction spanning a large range of non-zero read counts (Figure 6D). Intriguingly, two mice retained relatively high sgRNA diversity in late primary tumors (Figure 6D), likely due to dormant or slowly proliferating cells that remained in low numbers during tumor growth. Similar to the genome-wide library, large shifts in the sgRNA distribution exist between different sample types (KS test,  $p < 10^{-15}$  for pairwise comparisons between the cell, primary tumor, and lung metastases,  $p = 0.02$  between plasmid and cell) (Figure 6E).

In the validation minipool, the sgRNAs detected in the late primary tumors or the lungs of five different mice significantly overlap with each other (Figures S7E and S7F). The late primary tumors and lung metastases are dominated by a few sgRNAs (Figures 7A and S7G–S7I), suggesting that these sgRNAs outcompete others during tumor growth and metastasis. With the validation library, the sgRNA representations are highly correlated between late primary tumors and lung metastases (correlation,  $\rho = 0.55$  on average, F test,  $p < 0.01$ ,  $n = 5$ ) (Figure 7B). The late primary tumors and lung metastases have dozens of sgRNAs at moderate to high frequencies (Figures 7B and 7C). Several genes have multiple independent sgRNAs that are enriched in the lung over the primary tumor ( $\text{MPR} > 1$ ), such as *Nf2* (eight sgRNAs), *Pten* (four sgRNAs), *Trim72* (three sgRNAs), *Ube2g2* (three sgRNAs), *Ptges2* (two sgRNAs), and ATP-dependent DNA ligase IV (*Lig4*) (two sgRNAs) (Figures 7C and 7D). Two *Cdkn2a* sgRNAs were present in both late primary tumors and lung metastases in two mice, but with  $\text{MPR} < 1$ . *Fga*-, *Cryba4*-, *miR-152*-, and *miR-345*-targeting sgRNAs were not found at high frequency in either late primary tumors or lung metastases, suggesting that they are outcompeted by other loss-of-function mutations (such as *Nf2*), which agrees with the relatively reduced metastasis formation of these genes in the

(D) Inset: all sgRNAs found in individual lung lobes, ordered by the percent of lobes in which a particular sgRNA was among the significantly enriched ( $\text{FDR} < 0.2\%$ ) sgRNAs for that lobe. Only sgRNAs enriched in two or more lobes are shown. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

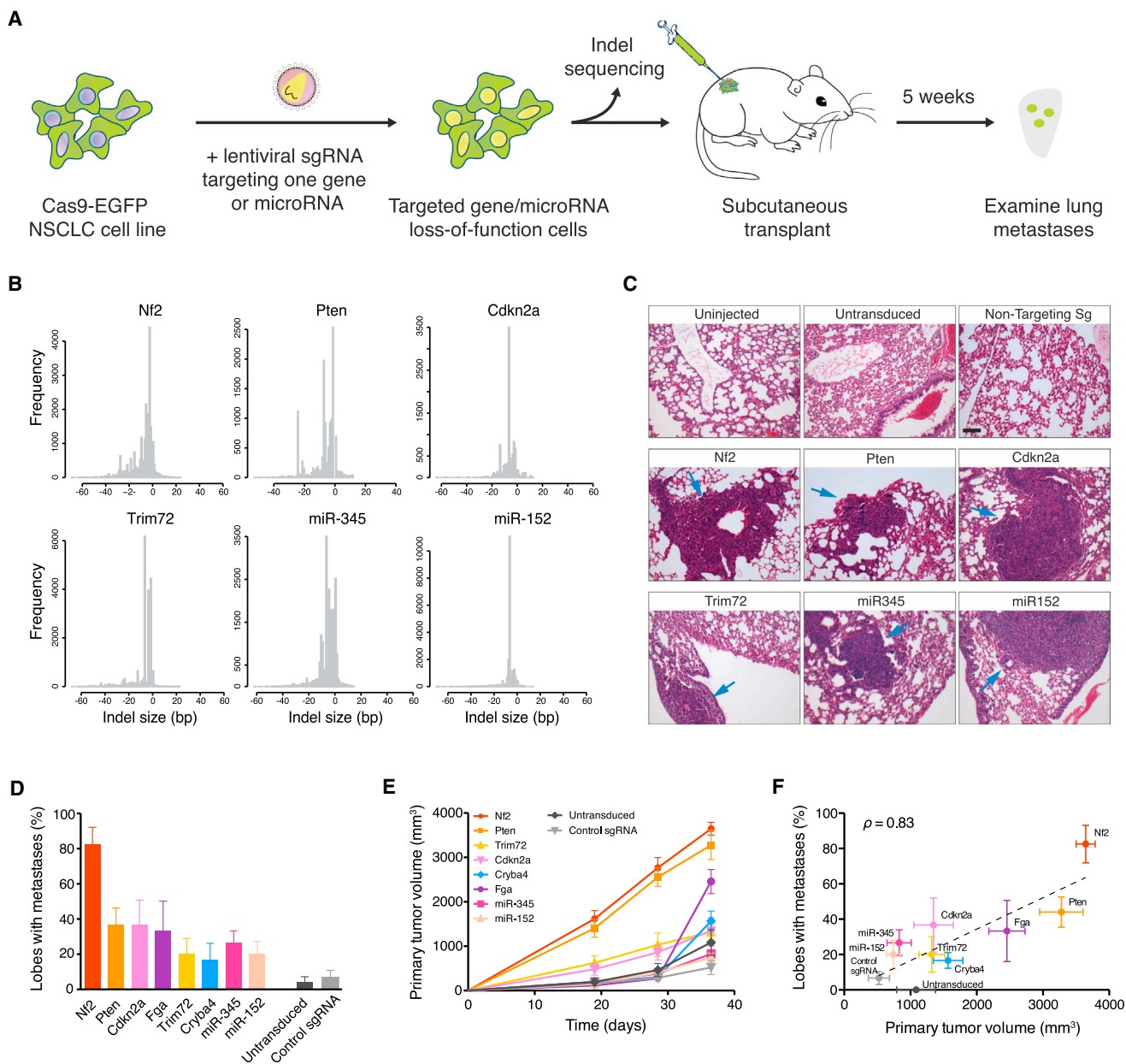
(E) Inset: all sgRNAs found in individual mice (averaged across three dissected lobes), ordered by the percent of mice in which a particular sgRNA was among the significantly enriched ( $\text{FDR} < 0.2\%$ ) sgRNAs for that mouse. Only sgRNAs enriched in two or more mice are shown. Main panel: enlargement and gene labels for sgRNAs at the top of the list from the inset (boxed region).

(F) Bottom: metastasis primary ratio (MPR) for the sgRNAs in mGeCKOa with enrichment in metastases over late tumors ( $\text{MPR} > 1$ ) observed in at least three mice. The sgRNAs are sorted by the number of mice in which the MPR for the sgRNA is greater than 1. Top: number of mice in which the MPR for this sgRNA is greater than 1. In both panels, individual sgRNAs are labeled by gene target.

(G) Number of genes with 0, 1, 2, or 3 significantly enriched ( $\text{FDR} < 0.2\%$  for at least one mouse) mGeCKOa sgRNAs in the lung metastases. For genes with 2 enriched sgRNAs, gene names are indicated in the colored bubble adjacent to the bar.

(H) Number of mice and percentage of mice in which each sgRNA was enriched in the lung metastases for all genes with multiple enriched sgRNAs.

See also Figures S4 and S5.



**Figure 5. Validation of Target Genes and MicroRNAs from mGeCKOa Screen Using Individual sgRNAs**

(A) Schematic representation of lentiviral transduction of Cas9-GFP KPD cells with single sgRNAs designed to target one gene or miR. After puromycin selection, the cell population was transplanted into *Nu/Nu* mice and also deep sequenced to examine the distribution of indels at the target site. After 5 weeks, the primary tumor and lungs were examined.

(B) Histograms of indel sizes at the genomic locus targeted by a representative sgRNA for each gene/miR after 3 days of puromycin selection. Indels from sgRNAs targeting the same gene were pooled (6 sgRNAs for each protein-coding gene; 4 sgRNAs for each miR).

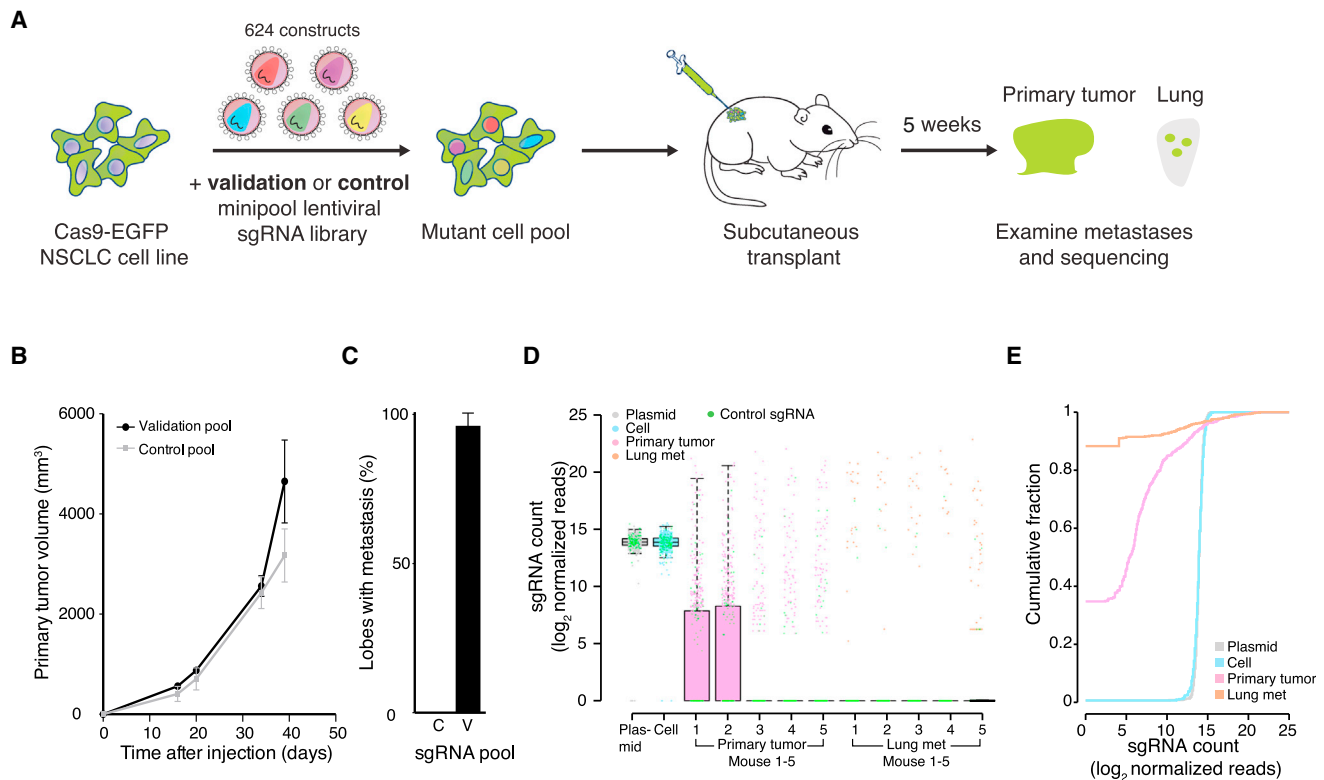
(C) Representative H&E staining of lung lobes from uninjected mice (n = 3 mice), mice transplanted with cells transduced with Cas9 only (n = 5), and mice transplanted with cells containing Cas9 and a single sgRNA (n = 6). Single sgRNAs are either control/non-targeting sgRNAs (n = 6 mice for control sgRNAs, 3 distinct control sgRNAs with 2 mice each) or targeting sgRNAs (n = 6 mice for each gene/miR target, 3 sgRNAs per target with 2 mice each). Blue arrows indicate lung metastases. Scale bar, 10  $\mu$ m.

(D) Percent of lung lobes with metastases after 6 weeks for the mice in (C). Error bars indicate SEM.

(E) Primary tumor growth curve of *Nu/Nu* mice transplanted with NSCLC cells transduced with Cas9 only (n = 5) or single sgRNAs (n = 6 mice per gene/miR target, 3 sgRNAs per target with 2 mice each; n = 6 mice for control sgRNAs, 3 control sgRNAs with 2 mice each). Error bars indicate SEM.

(F) Correlation between primary tumor volume and percent of lobes with metastases for each gene in (D) and (E). Error bars indicate SEM.

See also Figure S6.



**Figure 6. Tumor Evolution and Library Representation in Transplanted Cas9-GFP KPD Cells with Minipool Libraries**

(A) Schematic representation of the loss-of-function metastasis minipool screen. Briefly, Cas9-GFP KPD cells were transduced with either validation minipool (524 gene-targeting + 100 non-targeting sgRNAs) or control minipool (624 non-targeting sgRNAs). After puromycin selection, the cell pools were transplanted into *Nu/Nu* mice. After 5 weeks, validation minipool sgRNAs were sequenced from primary tumor and lung samples.

(B) Primary tumor growth curve of *Nu/Nu* mice transplanted with Cas9 vector + validation minipool cells ( $n = 5$  mice) or Cas9 + control minipool cells ( $n = 5$  mice). Error bars indicate SEM.

(C) Percent of lung lobes with metastases after 6 weeks for the mice in (B). C, control minipool; V, validation minipool. Error bars indicate SEM.

(D) Boxplot of the sgRNA normalized read counts for the plasmid library, cells before transplantation, primary tumors, and lung metastases using the validation minipool.

(E) Cumulative probability distribution of library sgRNAs in the validation plasmid pool, cells before transplantation, primary tumors, and lung metastases. Distributions of primary tumor and lung metastases are averaged across five mice.

See also Figure S7.

individual sgRNA validation. These results further validate several of the top hits from the primary screen, using either sgRNA dominance (e.g., *Nf2*, *Pten*, *Trim72*) or MPR (e.g., *Nf2*, *Trim72*, *Ube2g2*, *Ptges2*). This validation minipool reveals the dynamics of multiple competing mutants chosen from the primary screen hits and indicates that mutants with strong pro-growth effects tend to enhance metastasis (Figure 7E).

#### TCGA Gene Expression of Screen Hits in Human Lung Cancer

To assess the relevance of our mGeCKOa and validation minipool screen hits (genes targeted by sgRNAs enriched in lung metastases) to pathological metastasis in human cancer, we performed gene expression analysis of the human orthologs of these genes. We compared mRNA levels in metastatic compared to non-metastatic primary tumors in patient samples using TCGA mRNA sequencing data. We found that most (61%–75%) of these genes are downregulated in metastatic tumors in

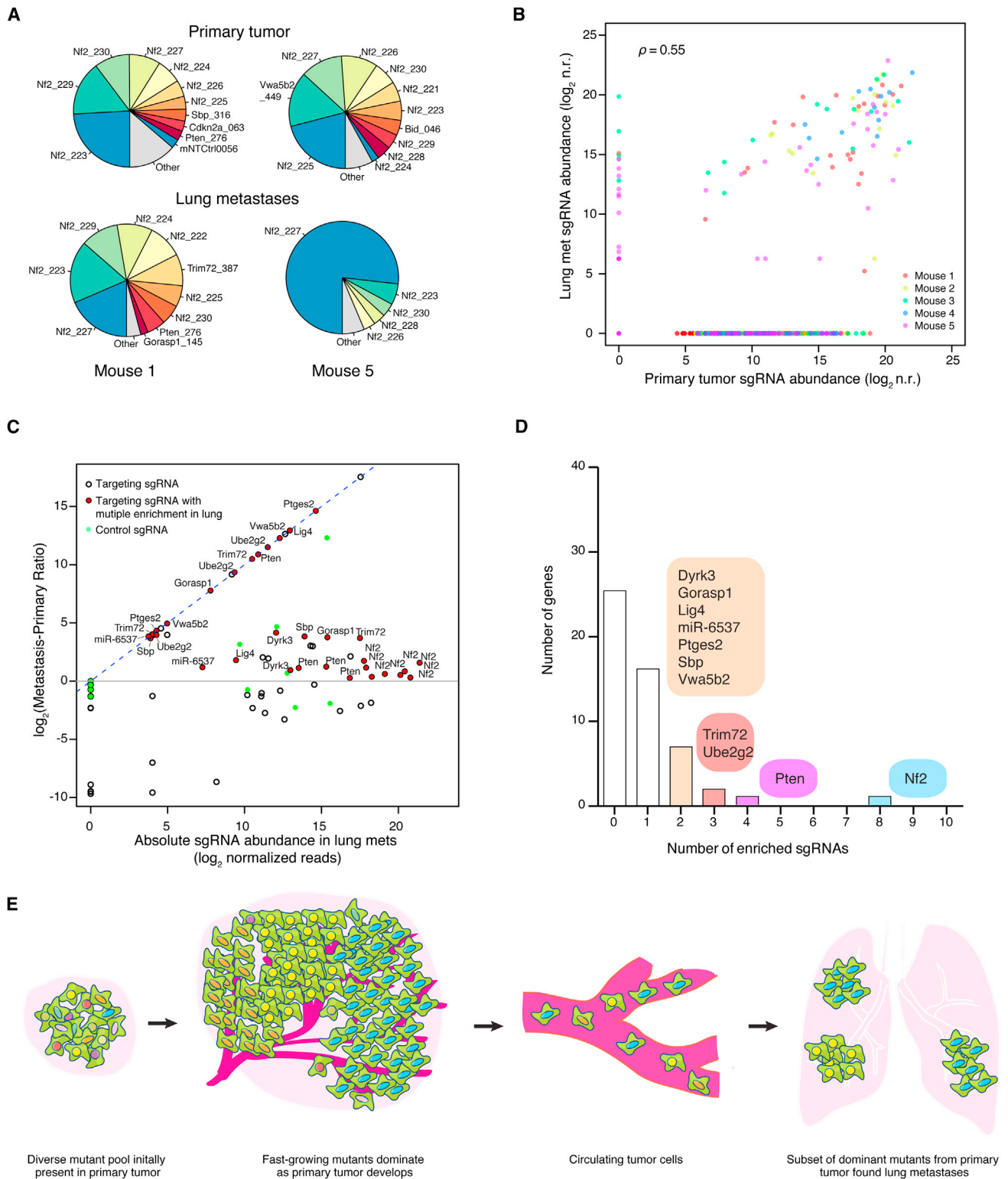
NSCLC patients (Figures S5D and S5E; Table S6). These data suggest that downregulation of these genes is selected for in metastatic tumors from patients.

#### DISCUSSION

##### Pooled Mutagenesis in a Metastasis Model

Distal metastases develop as primary tumors shed CTCs into the circulation, from which CTCs travel to the destination site, move out of the blood or lymphatic vessels, and initiate clonal growth (Valastyan and Weinberg, 2011; Vanharanta and Massagué, 2013; Weinberg, 2007). In this study, cancer cells transplanted into the flanks of mice form primary tumors in situ, and cells from this mass undergo the intravasation-circulation-extravasation-clonal growth cascade to form distal metastases (Francia et al., 2011). The initial lung cancer cell line has little capacity to form metastases; in contrast, after being mutagenized with the mGeCKOa genome-scale Cas9 knockout library, the cell





**Figure 7. Enriched sgRNAs from the Validation Minipool Screen in Primary Tumors and Lung Metastases**

(A) Pie charts of the most abundant sgRNAs in the primary tumor and the whole lung of two representative mice transplanted with validation minipool-transduced Cas9-GFP KPD cells. The area for each sgRNA corresponds to the fraction of total reads from the tissue (primary tumor or lung metastases) for the sgRNA. All sgRNAs with  $\geq 2\%$  of total reads are plotted individually.

(legend continued on next page)

population forms highly metastatic tumors. Thus, these mutations, acting in simple or complex pleiotropic ways, accelerate metastasis. In this model, the effect of mutations on metastasis strongly correlates with their abundance in late-stage primary tumors.

### sgRNA Dynamics during Tumor Evolution

The dynamics of the sgRNA population changed dramatically over the course of tumor development and metastasis, reflecting the selection and bottlenecks of cellular evolution *in vitro* and *in vivo*. After a week in culture, cells retained most of the sgRNAs present in the plasmid library, with decreases in sgRNAs targeting genes involved in fundamental cellular processes. The distribution of non-targeting control sgRNAs is almost identical to those targeting genes, suggesting that the selective pressure of *in vitro* culture alone does not radically alter sgRNA representation, similar to previous observations in human melanoma cells (Shalem et al., 2014).

In contrast, less than half of the sgRNAs survive in an early-stage primary tumor. This loss of representation occurs with both gene-targeting sgRNAs and non-targeting control sgRNAs, suggesting that random sampling influences sgRNA dynamics during the transplantation and tumor initiation processes, although we cannot exclude that some of the non-targeting sgRNAs might have detrimental or pro-growth effects. We also detected further dropout of genes involved in fundamental cellular processes in early tumor samples compared to cell samples. Thus, it is likely that the sgRNA dynamics are influenced by a combination of selection and random sampling during transplantation and tumor initiation.

As primary tumors grow, the mutant cells proliferate and compete as a pool. This creates strong selection for sgRNAs targeting anti-apoptotic genes and other tumor suppressors. The majority of the genetic diversity in early tumors is lost during the subsequent 4 weeks of primary tumor growth in mice. Accordingly, sequencing revealed a smaller set of dominant sgRNAs, usually on the order of hundreds to a few thousand per mouse. In addition, almost all of non-targeting sgRNAs are lost during primary tumor growth, which is consistent with selection for cells with special growth and survival properties. This observation is also consistent with earlier transplantation studies by Kerbel and colleagues using small pools of randomly mutagenized cells, which found that the majority of clonal variants detectable by Southern blot disappeared within 6 weeks of primary tumor growth, leaving one dominant clone (Korczak et al., 1988; Waghorne et al., 1988).

Each step toward metastasis has a bottleneck effect. In the lung metastases, we detected very few sgRNAs at high abun-

dance. As with the primary tumor, we found only a few non-targeting sgRNAs at low frequencies in metastases. Their presence could be due to unknown off-target effects of these sgRNAs, random shedding of CTCs in the primary tumor, or clustering together with other strongly selected CTCs during metastasis (Aceto et al., 2014).

### Relevance of Screen Hits to Human Cancer

Several of the genes enriched in late-stage primary tumors are associated with cancer, but their functions in tumor growth are poorly understood. For example, *Mgmt*, a gene with two enriched sgRNAs, is required for DNA repair and is thus crucial for genome stability (Tano et al., 1990). Mutation, silencing, or promoter methylation of *MGMT* is associated with primary glioblastomas (Jesien-Lewandowicz et al., 2009). *Med16*, another gene with two enriched sgRNAs, encodes a subunit of the mediator complex of transcription regulation, which has been recently implicated in cancer (Huang et al., 2012; Schiano et al., 2014).

We found that the genes that are significantly enriched in lung metastases largely overlap with those found in abundance in the late primary tumor. Several of these hits were validated *in vivo* using multiple individual sgRNAs, including *Nf2*, *Pten*, *Cdkn2a*, *Trim72*, *Fga*, *miR-152*, and *miR-345*. *Nf2*, *Pten*, and *Cdkn2a* are well-known tumor suppressor genes. Intriguingly, the *NF2* locus is mutated at only 1% frequency in primary tumors of human NSCLC patients (LUAD and/or LUSC) (Cancer Genome Atlas Research Network, 2012, 2014). *Nf2* mutant mice develop a range of highly metastatic tumors (McClatchey et al., 1998). It is possible that *NF2* mutations influence metastases to a greater degree than primary tumor growth, but this awaits metastasis genomics from patient samples. *Pten* mutations are also associated with advanced stages of tumor progression in a mouse model of lung cancer (McFadden et al., 2014), and *PTEN* was found to be mutated at 8% in adenocarcinoma patients (LUAD). *CDKN2A* has been shown to be often inactivated in lung cancer (Kaczmarczyk et al., 2012; Yokota et al., 2003). *Fga* encodes fibrinogen, an extracellular matrix protein involved in blood clot formation. *Fga* mutations have been found in various cancer types in TCGA (Lawrence et al. 2013), as well as circulating tumor cells (Lohr et al., 2014). *Trim72* is an E3 ubiquitin ligase, and its role in cancer metastasis is largely unknown. Studies have shown that *miR-152* and *miR-345* are associated with cancer and metastasis (Cheng et al., 2014; Tang et al., 2011). *FGF2* and *BAG3*, which promote metastasis, were predicted targets of *miR-152* and *miR-345*; thus, loss of these

(B) Scatterplot of normalized sgRNA read counts in primary tumor and lung metastases for all sgRNAs in the validation minipool for each mouse (different color dots indicate sgRNAs from different mice).  $\log_2$  n.r.,  $\log_2$  normalized reads.

(C)  $\log_2$  ratio of sgRNA abundance in the lung metastases over the primary tumor (MPR) plotted against the abundance in the lung metastases ( $n = 5$  mice per sgRNA). Green dots are the 100 control sgRNAs. Dots with black outlines are non-control sgRNAs that target genes or miRs. Red dots indicate non-control sgRNAs for which more than one sgRNA targeting the same gene/miR is enriched in the lung metastases over the primary tumor (i.e.,  $\log_2(\text{MPR}) > 0$ ) and are labeled with the gene/miR targeted. The lung-primary ratio is calculated for individual mice, and these quantities are averaged across mice.

(D) Number of genes with 0 to 10 significantly enriched validation minipool sgRNAs in lung metastases. For genes/miRs with 2 or more enriched sgRNAs, genes/miRs are categorized by how many sgRNAs targeting that gene/miRs are enriched, as indicated in the colored bubbles adjacent to each bar.

(E) Schematic illustration of tumor growth and metastasis in the library-transduced NSCLC transplant model. The initially diverse set of loss-of-function mutations in the subcutaneously transplanted pool is selected over time for mutations that promote growth of the primary tumor. A subset of these mutants also dominate lung metastases.

See also Figure S7.

microRNAs may lead to acceleration of metastases, likely due to de-repression of these genes (Cheng et al., 2014; Tang et al., 2011).

In our own analysis of TCGA samples from lung cancer patients, we observed downregulation of the human orthologs of the genes identified in the genome-wide and validation minipool screens at the mRNA level in metastatic tumors compared to non-metastatic tumors, suggesting that these genes may also be inactivated during pathological metastasis. Human orthologs of these genes are often found to be mutated in cancers. Moreover, these genes have been implicated in various pathways and biological processes in tumorigenesis and/or metastasis in human cancer (Tables S7A–S7C). However, most cancer sequencing studies involve samples from primary tumors of patients. In the clinic, metastases are rarely sampled. Future patient sequencing directly from metastases may further connect genes identified in the mouse model to those mutated or silenced in clinical metastases.

### Future In Vivo Functional Genomic Screens

Our study provides a roadmap for in vivo Cas9 screens, and future studies can take advantage of this model to explore other oncogenotypes, delivery methods, or metastasis target organs. Genome-scale CRISPR screening is feasible using a transplant model with virtually any cell line or genetic background (e.g., mutations in *EGFR*, *KRAS*, *ALK*, etc.), including a large repertoire of human cell lines from diverse cancer types (Barretina et al., 2012). Other cell delivery methods, such as intravenous injection or orthotopic transplantation, may help identify genes regulating extravasation and clonalization. Examining samples from other stages or sites, such as CTCs or metastases to other organs, can provide a more refined picture of tumor evolution.

In addition to these parameters, several aspects of the screen perturbations themselves can also be modified. Targeted drug therapies or immunotherapies can be applied in conjunction with the in vivo screening strategy to identify genes involved in acquired resistance. Other screening technologies, such as Cas9-mediated activation (Gilbert et al., 2014; Konermann et al., 2015), can identify metastasis-regulating factors that act in a gain-of-function manner. Activation screens that identify oncogenes, as well as dropout screens that identify genetic dependencies, may facilitate identification of novel therapeutic targets. Targeted subpool strategies can be used to reduce the library size and facilitate further confirmation of primary screens. In a customized library, genes can be chosen based on genomic analysis, pathways, or clinical relevance for focused screening libraries. Additionally, application of pooled sgRNA libraries using individually barcoded cells will allow quantitative assessment of the robustness and significance of each candidate hit and will enable analysis of the competitive dynamics among different perturbations. With these promising future directions and the results of our study, Cas9-based in vivo screening establishes a new platform for functional genomics discovery.

## EXPERIMENTAL PROCEDURES

### Generation of Cas9-GFP Expression Vector

A lentiviral vector, lenti-Cas9-NLS-FLAG-2A-EGFP (lentiCas9-EGFP), was generated by subcloning Cas9 into a lentiviral vector.

### Pooled Guide-Only Library Cloning and Viral Production

The Cas9-GFP KPD cell line was transduced at a MOI of ~0.4 with lentivirus produced from a genome-wide lentiviral mouse CRISPR knockout guide-only library (Sanjana et al., 2014) containing 67,405 sgRNAs (mGeCKOa, Addgene 1000000053) with at least 400-fold representation (cells per construct) in each infection replicate. A detailed viral production and infection protocol can be found in Extended Experimental Procedures.

### Animal Work Statement

All animal work was performed under the guidelines of the MIT Division of Comparative Medicine, with protocols (0411-040-14, 0414-024-17, 0911-098-11, 0911-098-14, and 0914-091-17) approved by the MIT Committee for Animal Care, and were consistent with the Guide for the Care and Use of Laboratory Animals, National Research Council, 1996 (institutional animal welfare assurance no. A-3125-01).

### Mice, Tumor Transplant, and Metastasis Analysis in the Primary Screen

Untransduced or mGeCKOa-transduced Cas9-GFP KPD cells were injected subcutaneously into the right side flank of *Nu/Nu* mice at  $3 \times 10^7$  cells per mouse. Transplanted primary tumor sizes were measured by caliper. At 6 weeks post-transplantation, mice were sacrificed and several organs (liver, lung, kidney, and spleen) were dissected for examination of metastases under a fluorescence stereoscope.

### Mouse Tissue Collection

Primary tumors and other organs were dissected manually. For molecular biology, tissues were flash frozen with liquid nitrogen and ground in 24-well polyethylene vials with metal beads in a GenoGrinder machine (OPS Diagnostics). Homogenized tissues were used for DNA/RNA/protein extractions using standard molecular biology protocols. Tissues for histology were then fixed in 4% formaldehyde or 10% formalin overnight, embedded in paraffin, and sectioned at 6  $\mu$ m with a microtome as described previously (Chen et al., 2014). Slices were subjected to H&E staining as described previously (Chen et al., 2014).

### Genomic DNA Extraction from Cells and Mouse Tissues

Genomic DNA from cells and tissues (primary tumors and lungs) was extracted using a homemade modified salt precipitation method similar to the Puregene (QIAGEN/Gentra) procedure. The sgRNA cassette was amplified and prepared for Illumina sequencing as described previously (Shalem et al., 2014). A detailed readout protocol can be found in Extended Experimental Procedures.

### Individual Gene and MicroRNA Validation

Six sgRNAs per protein-coding gene and four sgRNAs per microRNA gene were chosen for validation using individual sgRNAs (Table S4). For protein-coding genes, we cloned both the three sgRNAs from the mGeCKOa library and three additional sgRNAs to target each gene. For microRNAs, we used all four sgRNAs from the mGeCKOa library.

### Validation and Control Minipool Synthesis and In Vivo Transplantation

Validation and control minipools (Table S5) were synthesized using array oligonucleotide synthesis (CustomArray) and transduced at >1,000-fold representation in Cas9-GFP KPD cells. After 7 days in culture, Cas9-GFP KPD cells transduced with the validation minipool or control minipool were injected subcutaneously into the right side flank of *Nu/Nu* mice at  $3 \times 10^7$  cells per mouse with five replicate mice. After 5 weeks, mice were sacrificed, and primary tumors and lungs were dissected.

### ACCESSION NUMBERS

Genomic sequencing data have been deposited in the NCBI Sequence Read Archive under accession number PRJNA273894. Plasmids and pooled libraries have been deposited in Addgene (LentiCas9-EGFP: 63592, Metastasis Validation Minipool library: 63594, Mouse Non-targeting Control Minipool: 63595).

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, seven tables, and a dataset and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.038>.

## AUTHOR CONTRIBUTIONS

S.C., N.E.S., O.S., F.Z., and P.A.S. conceived and designed the study. S.C., N.E.S., and K.Z. performed all screening and validation experiments. S.C., N.E.S., O.S., and D.A.S. analyzed the data. K.L., J.S., R.W., and H.L. designed the CTC chip and performed CTC analysis. X.S. and J.Q.P. performed western blots. S.C., N.E.S., F.Z., and P.A.S. wrote the manuscript with the input from all authors. P.A.S. and F.Z. supervised the work.

## ACKNOWLEDGMENTS

We thank R. Weinberg and R. Kerbel for critically reading the manuscript; the entire P.A.S. lab and F.Z. lab, L. Cong, T. Kelly, X. Ni, M. Nobel, J. Boehm, A. Tsherniak, S. Levine, M. Cornwall-Brady, S. Maistrom, M. Jennings, E. Vasile, C. Whittaker, K. Cormier, R. Bronson, and colleagues at the Koch Institute, Broad Institute, McGovern Institute, and Department of Biology for technical assistance and/or discussion; the Swanson Biotechnology Center for technical support (Genomics, Animal Imaging and Preclinical Testing, Bioinformatics and Computing, Microscopy, Flow Cytometry, Microscopy, and Histology, in particular). This work is supported by grants to P.A.S., including a United States Public Health Service grant R01-CA133404 from the NIH, an MIT-Harvard Center for Cancer Nanotechnology Excellence Grant U54 CA151884 from the National Cancer Institute, a generous gift from the Marie D. and Pierre Casimir-Lambert Fund, an SKTech/MIT Initiative Grant from the Skolkovo Foundation, and the Koch Institute Support (core) grant P30-CA14051 from the National Cancer Institute. F.Z. is supported by the NIH through NIMH grant 5DP1-MH100706 and NIDDK grant 5R01-DK097768; a Waterman Award from the National Science Foundation; the Keck, New York Stem Cell, Damon Runyon, Searle Scholars, Merkin, and Vallee foundations; and Bob Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. S.C. is a Damon Runyon Cancer Research Fellow (DRG-2117-12) and also supported by the Dale Frey Award for Breakthrough Scientists. N.E.S. is supported by a Simons Center for the Social Brain Postdoctoral Fellowship and NIH NHGRI award K99-HG008171. O.S. is a fellow of the Klarman Cell Observatory. R.W. is supported by NIH grant U54 CA151884. D.A.S. is supported by a NSF Graduate Research Fellowship. H.L. is supported by Department of Defense grant OCRP W81XWH-14-1-0279. CRISPR reagents (plasmids and libraries) are available to the academic community through Addgene. F.Z. is a co-founder of Editas Medicine and a scientific advisor for Editas Medicine and Horizon Discovery.

Received: December 30, 2014

Revised: February 3, 2015

Accepted: February 18, 2015

Published: March 5, 2015

## REFERENCES

- Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* 158, 1110–1122.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extra-chromosomal origin. *Microbiology* 151, 2551–2561.
- Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519–525.
- Cancer Genome Atlas Research Network (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550.
- Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* 2, 401–404.
- Chen, S., Xue, Y., Wu, X., Le, C., Bhutkar, A., Bell, E.L., Zhang, F., Langer, R., and Sharp, P.A. (2014). Global microRNA depletion suppresses tumor angiogenesis. *Genes Dev.* 28, 1054–1067.
- Cheng, Z., Ma, R., Tan, W., and Zhang, L. (2014). MiR-152 suppresses the proliferation and invasion of NSCLC cells by inhibiting FGF2. *Exp. Mol. Med.* 46, e112.
- Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol.* 10, 726–737.
- Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 6091–6105.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607.
- Francia, G., Cruz-Munoz, W., Man, S., Xu, P., and Kerbel, R.S. (2011). Mouse models of advanced spontaneous metastasis for experimental therapeutics. *Nat. Rev. Cancer* 11, 135–141.
- Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, p11.
- Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67–71.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* 109, E2579–E2586.
- Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Heimann, R., and Hellman, S. (1998). Aging, progression, and phenotype in breast cancer. *J. Clin. Oncol.* 16, 2686–2692.
- Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* 31, 827–832.
- Huang, S., Hölzel, M., Knijnenburg, T., Schlicker, A., Roepman, P., McDermott, U., Garnett, M., Grenrum, W., Sun, C., Prahallad, A., et al. (2012). MED12 controls the response to multiple cancer drugs through regulation of TGF- $\beta$  receptor signaling. *Cell* 151, 937–950.
- Jesien-Lewandowicz, E., Jesionek-Kupnicka, D., Zawlik, I., Szybka, M., Kulczycka-Wojdala, D., Rieske, P., Sieruta, M., Jaskolski, D., Och, W., Skowronski,



- W., et al. (2009). High incidence of MGMT promoter methylation in primary glioblastomas without correlation with TP53 gene mutations. *Cancer Genet. Cytogenet.* 188, 77–82.
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- Kaczmarczyk, G., Lewandowski, R., Trautsohl, W., Ziolkowski, A., and Kozielski, J. (2012). Cytological examination of pleural cavity lavage accompanied by the study of gene promoter hypermethylation of p16 and O6-methylguanine-DNA-methyltransferase genes in diagnostics of non-small cell lung cancer metastatic changes into pleura. *Contemp Oncol (Pozn)* 16, 322–327.
- Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, Mdel.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* 32, 267–273.
- Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588.
- Korczak, B., Robson, I.B., Lamarche, C., Bernstein, A., and Kerbel, R.S. (1988). Genetic tagging of tumor cells with retrovirus vectors: clonal analysis of tumor growth and metastasis in vivo. *Mol. Cell. Biol.* 8, 3143–3149.
- Kumar, M.S., Pester, R.E., Chen, C.Y., Lane, K., Chin, C., Lu, J., Kirsch, D.G., Golub, T.R., and Jacks, T. (2009). Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev.* 23, 2700–2704.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Lohr, J.G., Adalsteinsson, V.A., Cibulskis, K., Choudhury, A.D., Rosenberg, M., Cruz-Gordillo, P., Francis, J.M., Zhang, C.Z., Shalek, A.K., Satija, R., et al. (2014). Whole-exome sequencing of circulating tumor cells provides a window into metastatic prostate cancer. *Nat. Biotechnol.* 32, 479–484.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- McClatchey, A.I., Saotome, I., Mercer, K., Crowley, D., Gusella, J.F., Bronson, R.T., and Jacks, T. (1998). Mice heterozygous for a mutation at the Nf2 tumor suppressor locus develop a range of highly metastatic tumors. *Genes Dev.* 12, 1121–1133.
- McFadden, D.G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.L., Cibulskis, K., Bhutkar, A., McKenna, A., Dooley, A., Vernon, A., et al. (2014). Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 156, 1298–1311.
- Nguyen, D.X., Bos, P.D., and Massagué, J. (2009). Metastasis: from dissemination to organ-specific colonization. *Nat. Rev. Cancer* 9, 274–284.
- Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* 14, 8096–8106.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* 11, 783–784.
- Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011). The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* 39, 9275–9282.
- Schiano, C., Casamassimi, A., Rienzo, M., de Nigris, F., Sommese, L., and Napoli, C. (2014). Involvement of Mediator complex in malignancy. *Biochim. Biophys. Acta* 1845, 66–83.
- Schramek, D., Sandoel, A., Segal, J.P., Beronja, S., Heller, E., Oristian, D., Reva, B., and Fuchs, E. (2014). Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science* 343, 309–313.
- Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., and Zhang, F. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87.
- Shao, D.D., Xue, W., Krall, E.B., Bhutkar, A., Piccioni, F., Wang, X., Schinzel, A.C., Sood, S., Rosenbluh, J., Kim, J.W., et al. (2014). KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell* 158, 171–184.
- Tang, J.T., Wang, J.L., Du, W., Hong, J., Zhao, S.L., Wang, Y.C., Xiong, H., Chen, H.M., and Fang, J.Y. (2011). MicroRNA 345, a methylation-sensitive microRNA is involved in cell proliferation and invasion in human colorectal cancer. *Carcinogenesis* 32, 1207–1215.
- Tano, K., Shiota, S., Collier, J., Foote, R.S., and Mitra, S. (1990). Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. *Proc. Natl. Acad. Sci. USA* 87, 686–690.
- Valastyan, S., and Weinberg, R.A. (2011). Tumor metastasis: molecular insights and evolving paradigms. *Cell* 147, 275–292.
- Vanharanta, S., and Massagué, J. (2013). Origins of metastatic traits. *Cancer Cell* 24, 410–421.
- Waghorne, C., Thomas, M., Lagarde, A., Kerbel, R.S., and Breitman, M.L. (1988). Genetic evidence for progressive selection and overgrowth of primary tumors by metastatic cell subpopulations. *Cancer Res.* 48, 6109–6114.
- Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343, 80–84.
- Weinberg, R.A. (2007). *The Biology of Cancer* (Garland Science).
- Yokota, J., Nishioka, M., Tani, M., and Kohno, T. (2003). Genetic alterations responsible for metastatic phenotypes of lung cancer cells. *Clin. Exp. Metastasis* 20, 189–193.
- Zender, L., Xue, W., Zuber, J., Semighini, C.P., Krasnitz, A., Ma, B., Zender, P., Kubicka, S., Luk, J.M., Schirmacher, P., et al. (2008). An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* 135, 852–864.
- Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509, 487–491.